Simon's 2-stage design + a bit of extensions

SSL Autumn Seminar 30.10.2014

Tiina Hakonen, Oncos Therapeutics

- ”Twenty years ago, in early 1990s, the term ”phase II trial design was practically synonymous with the Simon’s optimal and MINIMAX two stage (1989) –designs which have stood the test of time with their pracmatic trade-off between the need to stop a trial early for inefficacy if response rates were low and the likely overshoot of interim analysis in small trials.” Roger A’Hern (2014)

Outline

- 2-stage designs in Phase II oncology studies
- Simon's 2-stage design
- Sill and Yothers design
- Real life example from regulatory world

2-stage designs in Phase II oncology studies

- Phase I confirms the dose
- Phase II to estimate the degree of antitumor effect
- 2-stage design:
  - A small amount of patients are enrolled in the first stage, the enrollment of the another group of patients in stage 2 is conditional of the outcome of the first group
  - Typically activation of the second stage depends on the number of responses observe in the first stage
- ”Overshoot” of interim analysis in small studies?
  - Ethics: it is unethical to treat a full sample of patients with an inactive agent in setting involving life threatening disease such as cancer
  - But also, costs: one patient in phase II oncology trial ≈ 100,000 EUR

2-stage designs in Phase II oncology studies

- Phase I confirms the dose
- Phase II to estimate the degree of (*antitumor)* effect
- Generally single arm studies using short-term endpoints, typically tumor response, in limited number of patients
  - H_0: $p = p_0$ vs H_1: $p = p_A$
  - p_0 is the propability if true, it would mean that the regimen is not worth studying further (typically a value at or somewhat below the historical propability of response to standard treatment)
  - p_A is the propability if true, it would mean it would be important to identify the regimen as active and to continue studying it (typically a value somewhat above the historical propability of response to standard treatment)

2-stage designs in Phase II oncology studies

- Gehan 1961 (F)
- Fleming 1982 (F) (E)
- Simon 1987,1989 (F)
- Green & Dahlberg, 1992 (F)
- Heitjan, 1997 (”B”) (F) (E)
- Herndon, 1998 (F)
- Chen & Ng, 1998 (F)
- Chang et al, 1999 (F)
- Hanfelt et al, 1999 (F)
- Shuster, 2002 (F) (E)
- Tan & Machin, 2002 (B) (F)
- Case & Morgan, 2003 (F)
- Jung et al, 2004 (B) (F) (E)
- Lin and Shih, 2004 (F)
- Wang et al, 2005 (B) (F)
- Banerjee and Tsiatis, 2006 (B) (F) (E)
- Ye & Shyr, 2007 (F)
- Litwin el al, 2007 (F) (E)
- Wu & Shih, 2008 (F)
- Koyama & Chen, 2008 (F)
- Chi & Chen, 2008 (F) (E)
- Sambucini, 2008 (B) (F)
- ….

B = Bayesian, F = termination due to lack of activity, E = termination due to activity

The classical Simon's 2-stage design from 1989

- n_1 and n_2; number of patients studied in the first and second stage
- EN = expected sample size = $n_1 + (1 – PET)\ n_2$
- PET = probability of early termination after first stage
- Study will be terminated after first stage if r_1 or fewer responses are observed at first stage, this happens with probability PET = $B(r_1,p,n_1)$
- The drug (null hypothesis) will be rejected at the end of the second stage if r of fewer responses are observed.
- Probability of rejecting a drug after second stage with p is

$$B(r_1,p,n_1) + \sum b(x,p,n_1) \times B(r-x,p,n_2);\ x= r_1+1 \text{ to } \min(n_1,r)$$

- B denotes cumulative Binomial distribution, b Binomial probability mass function

Optimal and Minimax designs

Optimal

- Define parameters p_0, p_1, type I and II error rates α and β
- **Minimize the expected sample size when the response probability is p_0 i.e. early termination when the drug is not active**
- The optimization is taken over all values of n_1, n , r_1 and r (early acceptance of the drug not allowed)
- Fairly straight forward to build the optimization program in SAS, or copy the SAS-code from the WWW
- Cohen's paper covers $p_1 - p_0$ of 0.2 and 0.15

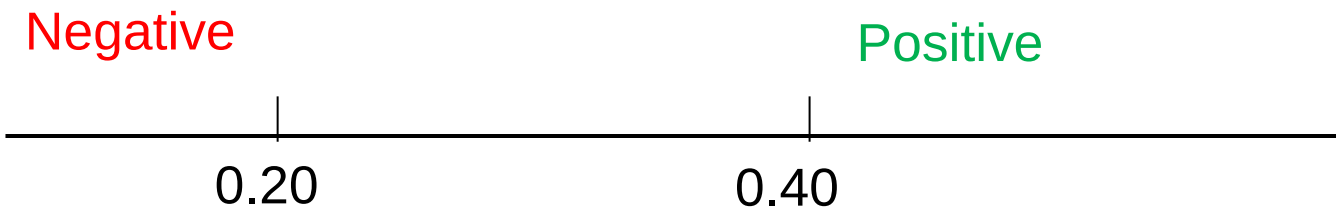
Minimax

- Optimal design does not necessarily minimize the maximum sample size
- On contrast, so called **Minimax** design have the smallest maximum sample size n (and secondly optimal)

Example

New promising treatment for platinum resistant ovarian cancer

Response rate in current standard of care 20%
Looking for 100% increase from 20% to 40%



H_0: p = 0.20
H_1: p > 0.20

Sample size, minimum n with
α = 0.05 and β = 0.20
P(X ≤ r | p=0.40) ≤ 0.20 and
P(X > r | p=0.20) ≤ 0.05
X = number of responses

Example

$p_0 = 0.20$, $p_A = 0.40$/ $\alpha = 0.05$ and $\beta = 0.20$

- One stage - exact-test minimum sample size n=35 and r = 11

- 2-stage optimal design

  $n_1 = 13$, $r_1 = 3$

  n = 43, r = 12

  $EN(p_0) = 20.6$, $PET(p_0) = 0.75$, $PET(p_1) = 0.17$, $EN(p_1) = 37.9$

- 2-stage minimax design

  $n_1 = 18$, $r_1 = 4$

  n = 33, r = 10 <= what! total sample size smaller than in one stage design

  $EN(p_0) = 22.3$, $PET(p_0) = 0.72$, $PET(p_1) = 0.09$, $EN(p_1) = 31.6$

Optimazing the sample size if $p = p_A$

By definition not Simon's 2-stage design

- Allows only early stopping for futility, but minimizes the expected sample size if $p = p_A$
- 2-stage optimal design in $p=p_A = 0.40$

  $n_1 = 18, r_1 = 4$

  $n = 33, r = 10$

  $EN(p_0) = 22.3, PET(p_0) = 0.72, PET(p_1) = 0.09, EN(p_1) = 31.6$

Same numbers as in minimax design
but it is not the same design [1]

[1] Mender, Thompson (2010) Contemp Clin Trials. 31(6): 562-578

Practical considerations

- Do I have to pause recruitment when number of patients needed for Stage I is reached? Yes
  - Herndon[1] (1998) proposes a hyprid design that allows continuation of recruitment while the results of the first stage are being analysed
- What if I have a few extra patients in stage I as we were not able to pause recruitement (patients were allready in screening phase!)
  - You may give some flexibility in your design, but define it in the protocol
    - Stage 1: recruit 13 -15 patient and define the rules for each option
    - Stage 2: recruit total of 43 – 45, and define the rules
    - So-called flexible designs (extension of the original) [2] [3]

  - [1] Herndon: Cont Clin Trials 1998, 19(5): 440-450
  - [2] Chen, Ng: Stat Med 1998; 17:2301-12
  - [3] Green, Dahlberg: Stat Med 1992; 11:853-62

Practical considerations

- Selection of end-point
  - typically response rate (complete or partial response) or disease control (complete & partial response and stable disease) measured at first imaging: after 2-3 treatment cycles
  - survival rate at 12-month? 3-month?
  - Biomarker
  - Need to define p_0, typically based on historical data, p_A (clinical relavant benefit $p_A - p_0$)
- Error rates
  - Weighting: what is more important 1) to discontinue uneffective treatment or 2) to move potentially effective treatment forward in development

Regulatory and sponsor error rates

	Optimal	Minmax	Optimal	Minmax
Stage I n_1 / r_1	13 / 3	18 / 4	17 / 3	9 / 0
Stage II n / r	43 / 12	33 / 10	37 / 10	36 / 10
PET(p_0)	0.75	0.72	0.55	0.13
PET(p_1)	0.17	0.09	0.05	0.01
EN(p_0)	20.6	22.3	26.0	32.4
EN(p_1)	37.9	31.6	36.1	35.4

$\alpha = 0.05$ and $\beta = 0.20$

$\alpha = 0.10$ and $\beta = 0.10$

One stage:
n=35 and r = 11

One stage:
n=36 and r = 10

Reporting the result

- Decision making based on responses seen
- May be supported by calculation of p-value, point estimate, confidence interval
  - P-value: $B(r_1,p_0,n_1) + \sum b(x, p_0,n_1) \times B(X-x,p_0,n_2)$; $x= r_1+1$ to $\min(n_1,X)$
    - X is the number of responses
  - Point estimate
    - ML: $\pi = X/n$ if study proceed to stage II else x_1/n_1
    - ML underestimates, however this tends to be corrected when if study proceed to Stage 2. A bias reduced estimator by Whitehead[1]
  - Confidence interval, be consistent with the hypothesis testing, i.e. one-sided $[\pi_L, 0)$. Method for example from Jennison[2]
- Critical, if the phase II is pivotal

[1] Whitehead (1986) Biometrika 73(3): 573-581
[2] Jennison & Turnbull (1999) Chapman and Hall

Back to our example

- 2-stage optimal design

  $n_1 = 13, r_1 = 3$

  $n = 43, r = 12$
- Continues to second stage as 4 responses observed at 1st stage
- At the end of stage 2 **14 responses** were observed
  - Null hypothesis rejected, ”drug effective”
  - ML-point estimate 14/43 = 0.326
  - p-value 0.0268
  - 90% CI: (0.217, *0.500*)

CI from http://data.vanderbilt.edu/~graywh/brew/twostage.html

Extensions

- Consideration of toxicity
- Inclusion of more than one treatment
- Addition of third stage
- Consideration of partial and complete responses separately
- Multiple strata
- Flexible designs

- Two binary co - endpoints according to Sill and Yothers

Sill & Yothers

- Sill MW, Rubinstein L, Litwin S and Yothers G: **A method for utilizing co-primary efficacy outcome measures to screen regimens for activity in two-stage Phase II clinical trials.** Clin Trials. 2012 Aug; 0(4): 385-395
- **http://www.gog.org/sdcstaff/mikesill/research/co-primary/Co-Primary%20Main%20Index.htm**
- Method for single arm study with dichotomous co-primary end points of efficacy, that has the ability to detect activity on either response measure with high probability when the drug is active on one or both measures, while at the same time rejecting the drug with high probability when there is little activity on both dimensions. The design enables early closure for futility and is **flexible with regard to attained accrual**.

Sill and Yothers in short

- A flexible, 2-stage trial design with an interim futility rule where interest is focused on detecting activity or either (or both) of two primary response variables
- Two end-points may represent diffent types of responses, however it is not sure if one or the other is more important/present
- Example end points:
  - Response rate - detect agents that are effective at selectively killing tumor cells
  - 3-month PFS – detect agents that can stabilize disease but are not necessarily effective at cell kill

Add second endpoint to the example

1. Response rate: $p_{10} = 0.20$, $p_{1A} = 0.40$
2. 3- month PFS: $p_{20} = 0.40$, $p_{2A} = 0.60$

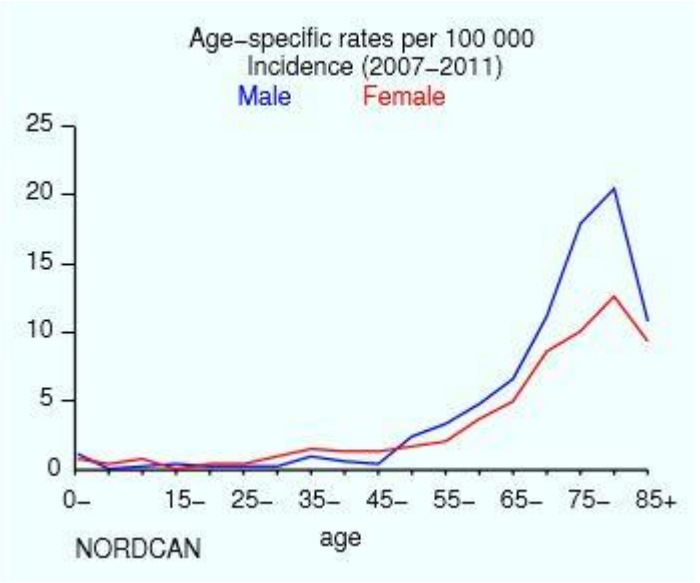
Flexibility not included in this setting

	One end point (RR) Optimal	1) Response rate	2) 3-month PFS
Stage I n_1 / r_1	13 / 3	13 / 3	13/ 5
Stage II n / r	43 / 12	25 / 9	25/14
PET(p_0)	0.75	0.43*	
PET(p_A)	0.17	Not done	
EN(p_0)	20.6	19.8	
EN(p_A)	37.9	Not done	

* PET is fairly low, might be good to look for designs with greater change to discontinue early due to futility. If 0.75 is set as lower limit of PET(p_0), the sample size increases to n_1=32 and n=42

Acute Myeloid Leukemia in Finland

- Number of new cases in per year (2007-2011): 75 males, 75 females
- Number of deaths: 65 males, 61 females
- Prevalence: 277 males, 376 females
- Relative survival (2009-2011)
  - 1-year 44% males, 45% females
  - 5-year 21% males, 23% females
- Disease of elderly
- The 1-year survival was below 40% in early 2000, i.e. has improved



A real life case study from year 2000...

- Gemtuzumab Zogamicin (G-Z) – an antibody-targeted chemotherapy agent in Acute Myeloid Leukemia (AML) patients
- Three single arm studies conducted:
  - 201-US 59 patients – PIVOTAL
  - 202-EU 25 patients
  - 203-WW 20 patients, above 60 years
  - At the time of NDA submission all studies continued recruitment
- 201 and 203 planned as Simon's design: $p_0 = 0.15$, $p_1 = 0.30$, $\alpha=0.10$ and $\beta=0.10$: $n_1 = 23/r_1=3$ n = 55/r=12. Primary endpoint = CR = Complete remission
- 202 was planned to provide more safety data

Story continues....

- 201: interim analysis was conducted when 23 were enrolled and completed part I of study. The target of 3+1 CRs was not met, however the study continued. And next analysis was done for NDA submission, when 59 patients were enrolled.
- 203: Interim was conducted for 20 patients
- Sponsor provided CRs from historical data (40%-54%) – WHAT the target was 30%

Summary of results

	Study 201	Study 202	Study 203	Pooled
Complete Remission (CR)	11/59 (19%)	4/25 (16%)	3/20 (15%)	18/104 (17%)
Morphologic Remission (MR)	9/59 (15%)	4/25 (16%)	1/20 (5%)	14/104 (13%)
Objective Remission, CR+MR (95% CI)	20/59 (34%) (22%-47%)	8/25 (32%) (15%-54%)	4/20 (20%) (6%-44%)	32/104 (31%) (22%-41%)
OS (months)				7.5 months (57 deaths)

Statistical reviewers comments...

- ”The CR results sugges that the treatment is inferior to the existing/publised literature results. But the NDA submission is not based on primary efficacy parameter CR, but based on the overall remission (OR), under the assumption that the CR and MR are the same.” OR = CR + MR
- ”Because of the open label, uncontrolled, non-randomized nature of the phase II trials presented in this NDA, no formal statistical testing or comparisons could be conducted. Therefore, any claims of ”improved efficacy” or ”no significant differences” need to be cautiously examined. The final recommendation should be based on clinical judgement.

And what happened?

- The treatment received Marketing authorization under Accelarated Approval pathway in 2000
- In 2004 a new study was initiated (as per requirement of the Accelarated approval), a comparative study: chemo vs. chemo + G-Z in the same population
- 2010 – no treatment benefit seen, more deaths in combination treatment
- The sponsor withdraw the drug for the market per FDA request

In 2013

- FDA granted **breakthrough designation** to volasertib for the treatment of patients with AML - not a marketing authorization!
- A Phase II study in patients with previously untreated AML ineligible for intensive therapy compared volasertib in combination with the established therapy of low-dose cytarabine (LDAC) versus LDAC alone. The primary endpoint for the Phase II study was **objective response**. Objective responses were observed in 31 percent of patients (13 of 42 patients) treated with the combination of volasertib* plus LDAC compared to 13.3 percent of the patients (6 of 45 patients) treated with LDAC alone (p = 0.0523).
- The sponsor has initiated a randomized phase III study (n=660)

Summary

- Simon's 2-stage design has well defended it's position as one of the most popular phase II study designs in oncology
- Has plenty of extensions
- Easy to understand by non-statisticians
- Has been been used successfully as pivotal study
- But lacks the control arm

Thank you!

Regulatory and sponsor error rates

- $p_0 = 0.20$, $p_A = 0.40$, $\alpha = 0.10$ and $\beta = 0.10$
- One stage
  - exact-test minimum sample size n=36 and r = 10
  - Normal approximation n=29 and r = 10
- 2-stage optimal design

  $n_1 = 17$, $r_1 = 3$

  n = 37, r = 10

  $EN(p_0) = 26.0$, $PET(p_0) = 0.55$, $PET(p_1) = 0.05$, $EN(p_1) = 36.1$
- 2-stage minimax design

  $n_1 = 9$, $r_1 = 0$

  n = 36, r = 10

  $EN(p_0) = 32.4$, $PET(p_0) = 0.13$, $PET(p_1) = 0.01$, $EN(p_1) = 35.4$