Outline

1. Motivation: Good Design and Analysis with Uncertainty About $\sigma^2$
2. Internal Pilots for a Gaussian Linear Model
3. Complications and Variations

Bibliography

---

1. Motivation: Good Design and Analysis with Uncertainty About $\sigma^2$

Must plan a study with uncertainty about nuisance parameters such as error variance, $\sigma^2$.

Want to avoid underpowered study and want to avoid overpowered study (time and cost savings).

1.1 Gaussian Error Linear Model Power Principles

For a fixed test and predictors, Gaussian linear model power depends on:

1) mean differences
2) variance
3) sample size.

---
1.2 Problem: Uncertainty About Nuisance Parameter $\sigma^2$

- t-Test Power for 3 Variances (Taylor and Muller 1995, 1996; Muller and Pasour, 1997)

1.3 Design Choices

1. Close your eyes and hope for the best.
2. Traditional external pilot study and design, two studies.
4. Adaptive designs, including internal pilot, group sequential, et al.

How do we provide accurate inference if we use data to adjust the design? Controversy surrounding some adaptive designs unintentionally implied guilt by association.

Design Choices (Coffey and Kairalla, 2008)

- Flexible designs
- Adaptive designs
- Sample size re-estimation
- Adaptive dose-response
- Seamless phase III/II designs
- Adaptive randomization
- Change other aspects of the trial (test statistic, primary endpoint, inclusion/exclusion criteria, dose, etc.)

Group Sequential Design Not Controversial

- Group sequential includes interim data analysis, which requires adjusted $\alpha$.
- Nearly all work for large samples (except Jennision and Turnbull, 1997).
- Assumes known $\sigma^2$, so not adaptive to it.

Fundamental design goal centered on allowing early stopping.

Opinion: in practice mostly a technique to allow peeking at the data with a nearly powerless test so modest cost to expected sample size.
### Internal Pilot Design Not Controversial

No interim data analysis, so no $\alpha$ cost for it (interim data analysis).
Small sample methods well developed for useful range of cases.
Can adjust sample size up or down for $\sigma^2$ for small $\alpha$ cost.


### Internal Pilot Steps for Univariate Gaussian Linear Model

1. **Plan**
   1.1 Choose design, test, target test size $\alpha_t$, power $P_t$, and means defining scientifically important effect.
   1.2 Use $\sigma_0^2$ to pick $n_0$ target total and $n_1 = \pi \cdot n_0$ sizes.
   1.3 Choose method for final $\hat{\sigma}^2$ and decision rule. Use GLUMIP.

2. **Conduct internal pilot**
   2.1 Collect $n_1$ observations, compute $\hat{\sigma}_1^2$.
   2.2 Power analysis finds $N_2 \geq 0$ observations to achieve $P_t$.

3. **Complete study**
   3.1 Collect $N_2$ observations and conduct (adjusted) analysis.

### 1.4 I Discourage Full Blinding

I recommend mean blinding but not mean and variance blinding.
Mean blinding: use design knowledge to compute model residual $\hat{\sigma}_1^2$ but remain blind to $\hat{\beta}$ ("noprint" option).
Total blinding: ignore design knowledge to compute model residual $\hat{\sigma}_1^2$.
Recommended by some (Gould, et al.).
Why should internal pilots be treated differently than group sequential?
Many references in the bibliography.
Could start with Waksman (2007) to illustrate many issues.

### 2. Internal Pilots for a Gaussian Linear Model

**Step 1.1 Plan a Fixed Sample Size Study**

**Example** Obstetrician Dr. Kirk Conrad plans to compare Cardiac Output (L/min) in 4 groups of women during pregnancy. Will record baseline (revised submission of P01 in review).
Fitting model as baseline + one compartment model in log space, corresponding to 7.4 maximum $= 4.9$ baseline $+ 2.5$ L/min. 50% change from baseline!
Fit a nonlinear model via transforming both sides with logarithm.
For $\mu_{0,A} = \log(\beta_0)$ test $H_0 : \mu_{0,A} = \mu_{0,B} = \mu_{0,C} = \mu_{0,D}$
Assuming $y = X\beta + \epsilon$ with $\epsilon \sim N(0, \sigma^2 I)$

<table>
<thead>
<tr>
<th>Design Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X = I_4 \otimes 1_m$</td>
<td>cell mean design with $m$ replicates</td>
</tr>
<tr>
<td>$\beta_{\text{plan}} = [0.92 \ 0.92 \ 0.92 \ 1.01]'$</td>
<td>Scientific important diff $\approx 10%$</td>
</tr>
<tr>
<td>$H_0 : C\beta = \theta_0$</td>
<td>Test by $C = [1_3 \ -I_3], \theta_0 = 0$</td>
</tr>
<tr>
<td>$\alpha_t$</td>
<td>target type 1 = 0.05/# primary outcomes</td>
</tr>
<tr>
<td>$P_t$</td>
<td>target power = 0.90</td>
</tr>
<tr>
<td>$\sigma_0^2$ here a SWAG</td>
<td>$(0.04)^2 \log(\text{L/min}) \Rightarrow n_0 = 20$</td>
</tr>
</tbody>
</table>

Step 1.2 Choose IP, Guess $\sigma_0^2 \Rightarrow n_0$ Target Total and $n_1 = \pi \cdot n_0$.

Yikes! If do interim power analysis at $n_1 = 12$ for $\pi = 0.50$
will have 3 women per group. What is your lowest choice?
Remember, we will fit a separate pharmacokinetic model to 7 values
(from echo cardiograms) of Cardiac Output, and test mean $\log(\hat{\beta}_0)$.
Consider recommending $n_1 = 16$ (4 per group)
for $\pi = 0.67$ (so 3 per group at interim power analysis).
A beautiful part of an internal pilot: we have done nothing different so far!
However, I have far less stress about the SWAG $\sigma_0^2$. 

* $y_w = \{y_0 + \beta_0[1 - \exp(-\beta_1 \cdot w)]\} \cdot e^{w}$
* $\beta_0 = 2.5 \text{L/min}$ gives the reference value

Means for 13 pregnant women (Robson, et al 1986)
Step 1.3 Choose IP Methods: Choice for Final $\hat{\sigma}^2$ and Decision Rule.

Choice # 1: how will $\sigma^2$ be estimated?
Using $\hat{\sigma}^2_{+}$ from the total sample can inflate type I error rate.

The amount of inflation varies with the parameter $\gamma = \frac{\sigma^2}{\sigma_0^2}$

Following plots are type I error rate and power
as a function of $\gamma = \frac{\sigma^2}{\sigma_0^2}$
for $\alpha_t = 0.0011$, $P_t = 0.90$, $n_1 = 10$, $n_0 = 20$,
$n_{+,\min} = 10$, $n_{+,\max} = \infty$

Choices include $\hat{\sigma}^2_{+}$ from first sample only
$\hat{\sigma}^2_{+}$, information added by second sample only
$\hat{\sigma}^2_{+}$, a weighted value to make unbiased
$\hat{\sigma}^2_{+}$, total sample with correction in small samples

I recommend a bounding method: use $\hat{\sigma}^2_{+}$ and critical value for $\alpha_s \leq \alpha_t$
for $\alpha_t = $ nominal type I error rate.
(Coffey and Muller 2001; Coffey, Kairalla, and Muller 2007)
Free SAS/IML code (GLUMIP version 2.0) at
http://www.jstatsoft.org/v28/i07 will give you $\alpha_s$
Step 1.3 Finding $\alpha_s \leq \alpha_t$ for Our Design

Adjusted Alpha $\alpha_s = 0.0338$ (20 lines of IML code)

PROC IML WORKSIZE=8000 SYMSIZE=8000;
%INCLUDE "&PROGPATH\GLUMIP20.IML" / NOSOURCE2;
ALPHAT = .05; POWERT = .90;
ESSENCEX = I(4); C = J(3,1,-1)||I(3);
BETA_PLN = {0.92, 0.92, 0.92, 1.01};
SIGMA0 = (.04)**2;
N1 = {12}; NPLUSMIN = N1; NPLUSMAX = 36;
RUN FINDADJ;
PRINT _FINDADJ[COLNAME=_FINDADJNM];

Step 2. Conduct Internal Pilot

Step 2.1 collect $n_1$ observations, compute $\hat{\sigma}_1^2$

Step 2.2 Power analysis finds $N_2 \geq 0$ observations needed to achieve $P_t$

Use standard power software, so very convenient.
The approach does create an "alignment error" with actual power, but modest. Improvements may be developed in future research.

Even here simulations a big, ugly, and very slow bear.

Step 3.2. Complete the Study

Step 3.1 Collect $N_2$ observations

Step 3.2 Conduct analysis with (adjusted) $\alpha_s$ if using bounding, using standard software.

Most other methods require some fiddling with variance estimates to piece together test statistics, but just simple programming (but be careful:)

REVIEW: Internal Pilot Steps for Univariate Gaussian Linear Model

1. Plan Choose design and analysis as usual.
Choose internal pilot size features.
Use GLUMIP for bounding method and refining design.

2. Internal pilot Collect $n_1$ observations, compute $\hat{\sigma}_1^2$
Power analysis $\Rightarrow N_2 \geq 0$ observations

3. Complete study Collect $N_2$ observations
Conduct analysis with adjusted $\alpha$
3. Complications and Future Work

1. Glossed over many issues around confidence intervals, stepdown tests. Answers depend on variance estimator. Does $\alpha_+$ suffice?
2. IP for other nuisance parameters? Some large sample approximations.
3. Random predictors, such as fractions in a blocking variable?
4. Non inferiority in small samples?! Caveat emptor, in small samples.
5. IP for large samples? Easy, should typically do it at no $\alpha$ cost. Common complaint centers on logistical barriers, funding. However, $N_+$ max most of fix, as in group sequential.

6. Research in Review and in Progress
   a) IP for some repeated measures; required theory and software coming
   b) small sample IP with 2 stage group sequential

Summary Review

1. Motivation: Good Design and Analysis with Uncertainty About $\sigma^2$
2. Internal Pilots for a Gaussian Linear Model:
   EASY TO DO.
   Not controversial.
   Small sample valid methods and free software available for many cases.
   Gives power insurance and cost protection.
3. Complications and Variations

Banish Power Uncertainty:
USE AN INTERNAL PILOT DESIGN!

I close by recommending the bibliography and online scholarly searches.
Brief bibliography for IP and power for Gaussian linear models; please see references in sources and use CIS.

Free SAS/IML code (GLUMIP version 2.0) for IP design and analysis at http://www.jstatsoft.org/v28/i07

***web site http://ehpr.ufl.edu/muller links to most of the following Muller articles and power software***

**General references for IP and power.**


**Power and IP articles with collaborators**


*****Articles by others about blinding in internal pilots********


