Within Subject Nested Clinical Trial Laboratory Data:

Sample Size, Components of Variance and Cost

Presented at Joint Statistical Meetings, Montreal, Quebec, 6 August 2013

Borko D Jovanovic, Hariharan Subramanian, Irene B Helenowski, Alfred W Rademaker, Angela J Fought, Hemant K Roy[1] and Vadim Backman[2]

Northwestern University Feinberg School of Medicine, Northwestern University[2] and Boston University[1]

borko@northwestern.edu

Abstract

Nesting of experimental factors is well established in statistical design literature relating to agricultural, environmental and engineering studies. It is less often discussed in biological and laboratory experiments stemming from the use of human bio-specimens, where sample size considerations are most often provided a priori ***on subject level***, and there is little advice regarding the needed number of units at lower levels. In this paper, motivated by an example from spectroscopic microscopy, we revisit the experimental nesting framework and discuss how variability, cost of sampling and sample size at lower levels may be coherently utilized. We show how the number of subjects may have to be adjusted to account for sampling decisions made at lower levels.

Key Words: Clinical Trials, Lung Cancer, Spectroscopic Microscopy, ANOVA

1. Introduction

In clinical trials, the number of sampling units to be studied on sub-subject level is often chosen according to existing laboratory folklore ("we always do three repeats"), and most often only the subject or group level data is reported and analyzed. The expected effect size is typically considered on a treatment group level; that is, as a result of an average across all existing levels: sub-cell level, cell level, tissues level, per human subject, and then per group.

Sample size calculations are then considered using an overall measure of variability observed in some past experiment, and considering the effect size that would make a clinical difference. Possible knowledge of variability at lower levels may be available, but is rarely included in the planning. This makes answer to the question '**how many items should be measured at lower levels**?' left to guessing or to institutional folklore. In this report we revisit the nesting framework and discuss how effect size and sample size at

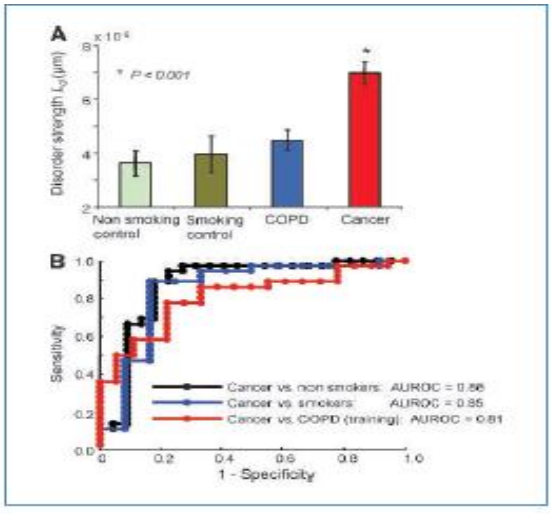
various levels may be thought of and coherently utilized. We also look into the possible orders of magnitude of cost of measurements at different levels of nesting.

2. Motivating Example: a Lung Cancer Study

The design of this observational study (Roy et al. 2010) involves collecting L_d measurements of "cell disarray" based on partial wave spectroscopic microscopy, from a population comprised of lung cancer patients, and three groups of controls: patients with COPD, smoking controls, and nonsmoking controls. Large values of L_d are in theory associated with disarray in cell structure and would suggest presence of cell stress eventually leading to development of cancer.

In the initial study measurements were recorded for each of 135 subjects, (cancer, COPD, smokers, non-smokers) with approximately 20-30 cells per subject, and within each cell, data were obtained from approximately 100,000-200,000 pixels per cell, each providing a measure of L_d. Such large number of pixels was provided by a machine which visually recorded the entire cell structure, as a part of a separate project. A summary of results is provided in Figure 1 below. Cancer patients have the largest average level of L_d, followed by COPD patients smoker-controls and finally by non-smoking controls. The ROC curves were formed and AUC(ROC) was observed to be in the 0.85 realm.

Figure 1. Summary of Roy et al. (2010) study results



Importantly, **the structure of measurement was such that pixels were nested within cells, cells were nested within subjects, while subjects were nested in several diagnosis groups** as in Figure 1. The underlying working hypothesis was that L_d levels sufficiently differ among patient groups so that a prediction rule may be developed and

tested prospectively in a future cancer screening trial, and a prophylactic prevention treatment could be applied to subjects at risk (subjects with high L_d).

The data was summarized and analyzed by averaging pixel intensity, providing values of the cell intensity, averaging over cells, thus providing a subject intensity and then, finally averaging subject intensities over groups of patients.

The question **we wished to address was: In planning a future clinical trial, given the observed variability at each level of data, what sample sizes should be selected at lower levels below subject level?**

Of course this question is related to specific components of variance which we look into next.

3. Components of variance and averages across sampling levels

We make some simple and predictable assumptions. Let X=x be the measurement at the pixel level, and assume it is independent from other observations on pixel level, and they all have common finite variance σ^2_{pix} . Then the averages across pixels in a cell have variance given by:

$$Var(\bar{x}_{pix}) = \frac{s^2_{pix}}{n_p}$$

Higher, on the cell level:

$$\sigma^2_{cell} = \sigma^2_{BetweenCells} + \sigma^2_{WithinCell} = s^2_{BetweenCells} + \frac{s^2_{pix}}{n_p}$$

Then the average *across cells* has variance:

$$Var(\bar{x}_{cell}) = \frac{s^2_{cell}}{n_c}$$

$$\frac{\sigma^2_{cell}}{n_c} = \frac{1}{n_c}\left(\sigma^2_{BetweenCells} + \frac{\sigma^2_{pix}}{n_p}\right)$$

On the subject level:

$$\sigma^2_{subject} = \sigma^2_{BetweenSubjects} + \sigma^2_{WithinSubject}$$

$$Var(\bar{x}_{subject}) = \frac{\sigma^2_{subject}}{n_s}$$
$$= \frac{1}{n_s}\left(\sigma^2_{BetweenSubjects} + \sigma^2_{WithinSubject}\right)$$
$$= \frac{1}{n_s}\left(\sigma^2_{BetweenSubjects} + \frac{1}{n_c}\left(\sigma^2_{BetweenCells} + \sigma^2_{WithinCell}\right)\right)$$
$$= \frac{1}{n_s}\left(\sigma^2_{BetweenSubjects} + \frac{1}{n_c}\left(\sigma^2_{BetweenCells} + \frac{1}{n_p}\sigma^2_{pix}\right)\right)$$

Finally, the last expression simplifies to a result we will find useful:

$$Var(\bar{X}_{subject}) = \frac{\sigma^2_{BetweenSubjects}}{n_s} + \frac{\sigma^2_{BetweenCells}}{n_s n_c} + \frac{\sigma^2_{BetweePixels}}{n_s n_c n_p}$$

Several things are worth noting here.

1. First, the **first summand in the formula above *is usually used to estimate the entire expression***.
2. Second, however one determines n_s, once it is determined other elements in the equation may be used to minimize, with appropriate constraints, the entire expression for variance.
3. Finally, one can study the trade-off among three sample sizes above, total variance, and total cost of the experiment.

In our motivational paper, Roy et al. (2010) observed:

$\sigma^2_{BetweenSubjects} = 0.308$, $\sigma^2_{BetweenCells} = 0.112$, and $\sigma^2_{pix} = 2.552$.

4. Sample size justification at lower levels as proportion of total variance of the mean

From established expression for variance of the overall mean across n_s subjects

$$Var(\bar{X}_{subject}) = \frac{\sigma^2_{BetweenSubjects}}{n_s} + \frac{\sigma^2_{BetweenCells}}{n_s n_c} + \frac{\sigma^2_{BetweePixels}}{n_s n_c n_p}$$

We can derive proportion of variability due to subjects and demand that sample sizes at lower levels guarantee that proportion of total variability due to lower levels is small, say 1% or smaller. This would translate to:

$$\frac{\frac{\sigma^2_{BetweenSubjects}}{n_s}}{\frac{\sigma^2_{BetweenSubjects}}{n_s}+\frac{\sigma^2_{BetweenCells}}{n_s n_c}+\frac{\sigma^2_{BetweePixels}}{n_s n_c n_p}} = 0.99$$

Notice that n_s cancels out from the left hand side, giving

$$\frac{\sigma^2_{BetweenSubjects}}{\sigma^2_{BetweenSubjects}+\frac{\sigma^2_{BetweenCells}}{n_c}+\frac{\sigma^2_{BetweePixels}}{n_c n_p}}$$

The inverse of which ***inflation ratio IR***:

$$\frac{\sigma^2_{BetweenSubjects}+\frac{\sigma^2_{BetweenCells}}{n_c}+\frac{\sigma^2_{BetweePixels}}{n_c n_p}}{\sigma^2_{BetweenSubjects}}$$

IR can be interpreted as *proportional increase in total variance due to lower (nested) levels*.

It has consequent effect on sample size calculation, as demonstrated below.

For example, consider the standard two sample t-test formula for sample size, with Type 1 error = 5%, alternative hypothesis is two-sided, and desired power is 80%:

$$n = \frac{2\sigma^2_{BetweenSubjects}(1.96+0.84)^2}{\Delta^2}$$

This sample size would have to be amended due to additional variability at lower levels.

The following table provides several values of the inflation ratio IR of the two sample sizes at lower levels.

Table 1. Percent increase in sample size needed for a future study given components of variance from past study in Roy et al. (2010)

n_c	3	5	10	50	10	100	20
n_p	3	5	10	10	100	100	150
IR%	208	80.8	23.8	4.77	8.93	0.89	4.19

Thus if we chose 3 observations per each lower level, we will need to increase subject level sample size by almost 210%, with 10 observations per lower level by about 24% and with 100 observations per level this becomes less than 1%.

Cost difference between the processing of a cell and processing of a pixel may add to deciding on optimality, and we discuss this next.

5. Sample Size Justification involving cost

Snedecor and Cochran (1967, p. 533) provide rationale for estimation of sample sizes on various levels using optimization via a cost function. Consider the cost of obtaining all of the samples on three levels as

$Cost = n_s \cos t_s + n_s n_c \cos t_c + n_s n_c n_p \cos t_p$ and variance of the mean across subjects

$$Var(\bar{X}_{subject}) = \frac{\sigma^2_{BetweenSubjects}}{n_s} + \frac{\sigma^2_{BetweenCells}}{n_s n_c} + \frac{\sigma^2_{BetweePixels}}{n_s n_c n_p} \text{ = Variance}$$

Then, using advanced calculus in derivation, the product

VC = Variance x Cost

can be minimized for

$$n_c = \sqrt{\frac{\cos t_s \sigma^2_{BetweenCells}}{\cos t_c \sigma^2_{BetweenSubjects}}} \quad \text{and} \quad n_p = \sqrt{\frac{\cos t_c \sigma^2_{pix}}{\cos t_p \sigma^2_{BetweenCells}}},$$

where $\mathbf{n_s}$ drops out from the equation.

In reality, it is either known beforehand, or found from the usual sample size considerations on subject level.

To verify these expressions for our data, we use the following information:

$\sigma^2_{BetweenSubjects} = 0.308$ $\sigma^2_{BetweenCells} = 0.112$ $\sigma^2_{pix} = 2.552$.

Note that variance at the lowest level is an order of magnitude larger than those at upper levels. This in principle **does not have to be the case** and reversed order of magnitude is possible as well.

We take an estimate that **cost per subject = $1,000, cost per cell = $1, cost per pixel = $0.001.** Simple application of formulas above provides:

$$n_p = \sqrt{\frac{1 \times 2.552}{0.001 \times 0.112}} = 150.94$$

$$n_c = \sqrt{\frac{1000 \times 0.112}{1 \times 0.308}} = 19.01$$

When these two **values are used in Table 1, as 150 and 20 approximately, we see that the total sample size on subject level has to be increased by about 4.2%.**

If the total sample size previously planned is 100, the adjusted sample size would be about 104 to have similar power as planned. This would translate into $4,000 additional cost if approximate cost per subject is $1,000, for a total of $104,000 for subject recruitment. For lab work we have 20x$1 + 150x$0.001 =$ 20.15 per subject or 104 x $20.15 = $2,095.60 for all subjects, for **the grand total cost of the trial of $106,095.60**. This of course assumes the trial drug or treatment is paid for from other resources or provided free of charge.

6. Discussion

Snedecor and Cochran (1967), provide an example of three stage sampling of **turnip green plants**: the first stage is plants, second stage is leaves within plans, the third stage is determinations within one leaf. Underwood (1997) provides an example of nested sampling via **orchards, trees, branches and twigs**. Sokal and Rohlf (2012) provide example of experiment involving drugs, **rats, rat livers and readings within livers**. Quinn and Keough (2002) provide an example of effect of **grazing of sea urchins** on percentage cover of filamentous algae. All these examples essentially provide the same solution to the questions raised in this article.

If total available cost of the experiment is provided, sample size on the subject level can be calculated to fit the cost constraints. In clinical trials however, one usually starts with the sample size on subject level, and not the total cost allowable for the trial. Laboratory or 'pathology' costs are calculated separately and are often unknown.

We suggest taking the following steps.

1. Get an estimate of variability on each sampling level.
2. Calculate sample size **on subject level first**, obtaining $\mathbf{n_s}$.
3. Find an optimal combination of sample sizes on lower levels, following arguments and methods provided in this paper.
4. Increase $\mathbf{n_s}$ as needed to achieve previously planned power.

References

1. Hemant K. Roy, Hariharan Subramanian, Dhwanil Damania, Thomas A. Hensing, William N. Rom,Harvey I. Pass, Daniel Ray, Jeremy D. Rogers, Andrej Bogojevic, Maitri Shah,Tomasz Kuzniar, Prabhakar Pradhan, and Vadim Backman: **Optical Detection of Buccal Epithelial Nanoarchitectural Alterations in Patients Harboring Lung Cancer: Implications for Screening.** *Cancer Res; 70(20) October 15, 2010.*
2. Quinn GP, Keough MJ**. Experimental Design and Data Analysis for Biologists**. Cambridge University Press. New York. 2002.
3. Snedecor GW, Cochran WG. **Statistical Methods, 6th Edition**. Iowa State University Press. Ames, Iowa 1967.
4. Sokal RR, Rohlf FJ. Biometry. **The Principles and Practice of Statistics in Biological Research**. Fourth Edition. WH Freeman and Co. New York. 2012.
5. Underwood AJ. **Experiments in Ecology. Their logical design and interpretation using analysis of variance**. Cambridge University Press. Cambridge, UK. 1997.