Open Access   CrossMark

Clinical trial laboratory data nested with in subject: components of variance, sample size and cost

Abstract

Nesting of experimental factors is well established in statistical design literature related to agricultural, environmental and engineering studies. It is perhaps not sufficiently discussed in biological and laboratory experiments stemming from the use of human bio-specimens, where sample size considerations are often provided a priori *on subject level*, but there is little advice regarding the *needed number of units at lower levels*. Motivated by an example from spectroscopic microscopy and lung cancer, we revisit the experimental nesting frame work 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 inadequate sampling decisions made at lower levels.

Keywords: clinical trials, sample size, lung cancer, spectroscopic microscopy, ANOVA

Borko D Jovanovic,[1] Hariharan Subramanian,[2] Irene B Helenowski,[1] Hemant K Roy,[1] Vadim Backman[2]
[1]Northwestern University, Feinberg School of Medicine, USA
[2]Boston University, USA

Correspondence: Borko D Jovanovic, Feinberg School of Medicine, Northwestern University Department of Preventive Medicine, USA, Email borko@northwestern.edu

Introduction

In randomized clinical trials, the sample size (i.e. the number of subjects planned to be used) is carefully scrutinized, studied in statistics courses, and advised in this realm.[1] Quite contrary to that, the number of sampling units to be studied on *sub-subject level* is often ignored or chosen according to existing laboratory folklore: e.g. "we always do three repeats". Most often only the subject and group level data are reported and considered in sample size calculations. The expected *effect size* is typically considered on a treatment group level as a result of an average or summary across all existing levels: sub-cell level, cell level, tissues level, per human subject, and per treatment group. Sample size calculations are then based on overall measure of variability considering the putative effect size that would make a clinically important difference. *Possible knowledge of variability at lower nested levels* may be available, but is rarely included in the planning of a trial. This makes the answer to the question 'how many items should be measured at lower levels, left to budgetary limitations. In this paper, we revisit the nesting framework and discuss how effect size and sample size at various levels may be used in sample size calculations.

Motivating example: a lung cancer study

An observational study Roy et al.,[2] was used as a template for in preparation for designing a randomized clinical trial example. Particularly, it involved collecting L_d measurements of "disorder strength of cell nano architecture" in saliva swab samples, based on partial wave spectroscopic microscopy. The 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 nano architecture* and suggest presence of cell stress, potentially 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.
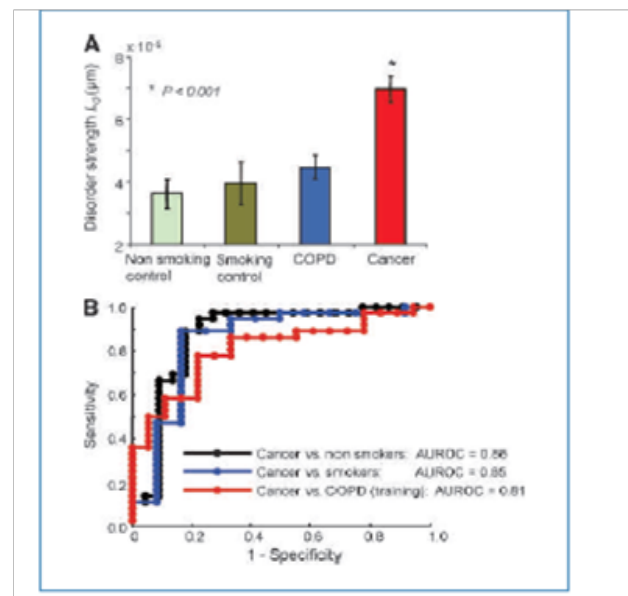

Figure 1 Summary of Roy et al. (2010) study results.

Importantly, measurements were structured 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 cancer and control groups so that a prediction rule may be developed and tested prospectively to detect yet undetected cancer cases. Alternatively a prophylactic prevention treatment could be applied to subjects at risk, subjects with high L_d, so that such measure of cell disarray would be brought to normal levels.

The original data were 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 means of a "COPD smoker" group and the "COPD only" group were 4.8 +/- 2.1 and 4.0 +/- 2.3, respectively. In order to distinguish between the 'high' and 'decreased' level of cell disarray, it was felt that a decrease of 25% over the level seen in the 'high' group would adequately deem an intervention aimed at reducing L_d as effective. Thus, using the standard, two sided, two sample t-test formulas for sample size, with Type 1 error = 95%, different variances and a power of 80%, for each group, one would need:

$$n = (2.1^2 + 2.3^2)(1.96 + 0.84)^2 / (4.8 - 4.0)^2 = 119$$

Subjects per group.

The next question is: what sample sizes should be selected at lower levels below subject level? This question is related to the specific components of variance which we look into next.

Components of variance and averages across sampling levels

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

$$Var\left(\overline{x}_{pix}\right) = \frac{\sigma^2_{pix}}{n_p} \quad (1)$$

And higher, on the cell level:

$$\sigma^2_{cell} = \sigma^2_{Betweencells} + \sigma^2_{WithinCell} = \sigma^2_{Betweencells} + \frac{\sigma^2_{pix}}{n_p} \qquad (2)$$

Then the average *across cells* has variance:

$$Var\left(\overline{x}_{cell}\right) = \frac{\sigma^2_{cell}}{n_c}$$

$$Var(\overline{x}_{subject}) = \frac{\sigma^2_{subject}}{n_s} \qquad (3)$$

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

$$Var(\overline{X}_{subject}) = \frac{\sigma^2_{BetweenSubjects}}{n_s} + \frac{\sigma^2_{BetweenCells}}{n_s n_c} + \frac{\sigma^2_{BetweePixels}}{n_s n_c n_p} \qquad (4)$$

Several things are worth noting here.

First, the first summand in the formula above *is usually used to estimate the entire expression*.

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.

Finally, one can study the trade-off among three sample sizes above, total variance, and total cost of the experiment.

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, given in equation (4).

We can derive proportion of variability due to subjects so 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:

$$\frac{\dfrac{\sigma^2_{BetweenSubjects}}{n_s} + \dfrac{\sigma^2_{BetweenCells}}{n_s n_c} + \dfrac{\sigma^2_{BetweenPixels}}{n_s n_c n_p}}{\dfrac{\sigma^2_{BetweenSubjects}}{n_s}} < 0.01 \qquad (5)$$

Notice that n_s cancel out from the left hand side, giving the *inflation ratio IR*:

$$\frac{\sigma^2_{Betweensubjects} + \dfrac{\sigma^2_{Betweensubjects}}{n_c} + \dfrac{\sigma^2_{BetweenPixels}}{n_c n_p}}{\sigma^2_{Betweensubjects}} \qquad (6)$$

IR can be interpreted as the *proportional increase in total variance due to lower (nested) levels*, and should remain low.

Using $\sigma^2_{BetweenSubjects} = 0.308$, $\sigma^2_{BetweenCells} = 0.112$, and $\sigma^2_{pix} = 2.552$, observed by Roy et al.,[2] we numerically compute:

IR = (0.308 + 0.112/n_c+2.552/$n_c n_p$)/0.308.

For various values of the two unknown sample sizes we can compute the inflation ratio (IR), the following table provides several values of IR of the two sample sizes at lower levels.

For example, if we chose 3 observations per each lower level, we will need to increase subject level sample size by 208%, or from n=119 to n*=250. With 10 observations per lower level we need to increase n by 23.8% or to n*=148, and with 100 observations per level this becomes less than 1%, a very tolerable increase to from n= 119 to n*=120. Cost difference between the processing of a cell and processing of a pixel may add to deciding on optimality discussed in the next section.

Sample size justification involving cost

Snedecor and Cochran[3] 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 $Cost = n_s cost_s + n_s n_c cost_s + n_s n_c n_p cost_p$, along with equation (4).

Then, using advanced calculus in derivation, the product:

VC = Variance x Cost   (7)

This can be minimized for:

$$n_c = \sqrt{\frac{\cos t_s \, \sigma^2_{BetweenCells}}{\cos t_c \, \sigma^2_{BetweenSubjects}}} \text{ and } n_p = \sqrt{\frac{\cos t_c \, \sigma^2_{pix}}{\cos t_p \, \sigma^2_{BetweenCells}}}, \quad (8)$$

Where n_s drops out from the equation: w

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

To verify these expressions for our data, we use:

Citation: Jovanovic BD, Subramanian H, Helenowski IB, et al. Clinical trial laboratory data nested with in subject: components of variance, sample size and cost. *Biom Biostat Int J.* 2015;2(3):81‒83. DOI: 10.15406/bbij.2015.02.00029

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

cost estimates provided below.

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

$$n_p = \sqrt{\frac{1 \times 2.552}{0.001 \times 0.112}} = 150.94 \qquad (9)$$

$$n_c = \sqrt{\frac{1000 \times 0.112}{1 \times 0.308}} = 19.01 \qquad (10)$$

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%.

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

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

If the total sample size previously planned is n=119, the adjusted sample size would be about n*=124 to have similar power. This would translate into a \$5,000 additional cost if approximate cost per subject is \$1,000, for a total of \$129,000 for subject recruitment. For lab work we have 20x\$1 + 150x\$0.001 =\$ 20.15 per subject or 124 x \$20.15 = \$2,498.6 for all subjects, for the grand total cost of the trial of \$131,498.60, assuming the trial drug or treatment is paid for from other resources.

One level of nesting only

In the context of the study described so far, we have cells nested in subjects and pixels nested in cells. Suppose now that pixel level does not exist but that an observation is made on each cell by some other means or some other technology. Then similar formulas follow and are applicable, as presented below.

$$Var(\bar{X}_{subject}) = \frac{\sigma^2_{BetweenSubjects}}{n_s} + \frac{\sigma^2_{BetweenCells}}{n_s n_c} \qquad (11)$$

The simplified expressions for cost still apply:

$$Cost = n_s cost_s + n_s n_c cost_s \text{ and the sample size on lower level,}$$
conditional on sample size on higher level is

$$n_c = \sqrt{\frac{\cos t_s \sigma^2_{BetweenCells}}{\cos t_c \sigma^2_{BetweenSubjects}}} . \qquad (12)$$

Discussion

The effect of nesting on experimental design has been a topic of interest in a vast range of equations in previous literature. Sokal and Rohlf[4] provide example of an experiment involving drugs, rats, rat livers and readings within livers. Quinn and Keough[5] provide an example of effect of grazing of sea urchins on percentage cover of filamentous algae. Snedecor and Cochran,[3] provide an example of three stage sampling of turnip green plants: the first stage is plants, second stage is leaves within plans, and the third stage is determinations within one leaf. Underwood[6] provides an example of nested sampling via orchards, trees, branches and twigs. 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 then suggest that in designing a trial, one should first get an estimate of variability on each sampling level and calculate sample size on subject level first, obtaining n_s. Next we recommend finding an optimal combination of sample sizes on lower levels, following arguments and methods provided in this paper. Finally, we should increase n_s as needed to achieve previously planned power.

Conclusion

We have exemplified this method using the lung cancer study in Roy et al. 2010, calculating the sample at each level and cost required to deem a difference between groups as statistically significant. Our resulting power analyses give feasible sample size and cost estimates compliant with our study design.

Acknowledgement

Conflict of interest

The authors declare that they have no financial or non-financial competing interests.

References

1. Julious SA. Sample Sizes for Clinical Trials. CRC Press, USA 2009.

2. Hemant KR, Hariharan S, Damania D, et al. Optical Detection of Buccal Epithelial Nanoarchitectural Alterations in Patients Harboring Lung Cancer: Implications for Screening. *Cancer Research*. 2010;70(20):7748–7754.

3. Snedecor GW, Cochran WG. *Statistical Methods*, Sixth Edition. Iowa State University Press. Ames, IA, USA, 1967.

4. Sokal RR, Rohlf FJ. *The Principles and Practice of Statistics in Biological Research*. Fourth Edition. WH Freeman and Co. USA, 2012.

5. Quinn GP, Keough MJ. *Experimental Design and Data Analysis for Biologists*. Cambridge University Press. USA. 2002.

6. Underwood AJ. Experiments in Ecology. *Their logical design and interpretation using analysis of variance*. Cambridge University Press. Cambridge, UK, 1997.