BASELINE ADJUSTMENT FOR STATISTICAL EFFICIENCY ON CLINICAL CONTROLLED TRIAL

Loh SY
Department of Rehabilitation , Faculty of Medicine, University of Malaya, Kuala Lumpur

ABSTRACT

In a clinical controlled trial involving repeated measures of continuous outcomes such as quality of life, distress, pain, activity level at baseline and after treatment, the possibilities of analyzing these outcomes can be numerous with quite varied findings. This paper examined four methods of statistical analysis using data from an outcome study of a clinical controlled trial to contrast the statistical power on those with baseline adjustment. In this study, data from a CCT with women with breast cancer were utilized. The experiment (n=67) and control (n=74) were about equal ratio. Four method of analysis were utilized, two using ANOVA for repeated measures and two using ANCOVA. The multivariate between subjects of the combined dependents variables and the univariate between subjects test were examined to make a judgement of the statistical power of each method. The results showed that ANCOVA has the highest statistical power. ANOVA using raw data is the least power and is the worst method with no evidence of an intervention effect even when the treatment by time interaction is statistically significant. In conclusion, ANOVA using raw data is the worst method with the least power whilst ANCOVA using baseline as covariate has the highest statistical power to detect a treatment effect other than method. The second best method as shown in this study was in using change scores of the repeated measures. *(JUMMEC 2009; 12 (1): 31-34)*

KEYWORDS: *ANCOVA, method, repeated measures, statistical power, research design*

Introduction

A common research design in randomized controlled trails is the use of baseline and repeated measures to study the changes as a result of intervention over a time period (1, 2, 3, 4). However, repeated measures design has its pro and cons. One major advantage of repeated measure design is the ability to control for the potential influence of individual differences (4, 5) resulting in differences observed amongst treatment conditions was more likely to reflect treatment effects and not the variability between the subjects. The disadvantages include more tedious measurement and the need to guard over practice effect or carry over. This simply means the subjects' increasing adept with the tasks and not that intervention has improved performance, and is particularly true for experimental study involving motor tasks.

As baseline measurement taken at the beginning of a study helps improve the research design by removing some of the variation in the data, there is a crucial need to ensure covariates was measured prior to intervention to avoid the scores on the covariates being influence by treatment(6). Consequently, these baselines must be considered carefully in analysis.

Methodology

In a clinical controlled trial conducted for women with breast cancer, half of the patients were allocated to receive an intervention and half to be in a control group. All patients were repeatedly measured for the same outcomes at the beginning of the study, at post intervention at 4th week and at the 8th week follow up. Data from the database of the SAMA (Staying abreast, Moving ahead) project were utilized for this paper. This study found that there were at

Correspondance:
Loh Siew Yim
Department of Rehabilitation
Faculty of Medicine, University of Malaya
50630 Kuala Lumpur
Email: syloh@um.edu.my

least four possibilities for analyzing the data with the baseline values; i) Analysis of Variance (ANOVA with raw scores; ii) ANOVA with change scores, iii) ANCOVA with baseline as covariate. All methods were analysed using The Statistical Package for the Social Sciences (SPSS) version 14. The General Linear Model's repeated measure ANOVA were used on the four methods specified below:

i. Repeated measures ANOVA using raw score: Treating the baseline and subsequent measurements as repeated measures in a repeated measures analysis of variance ANOVA)

ii. Repeated measures ANOVA using change scores: compute a change score (simply the difference between the second measurement and the first) and use that change score as the unit of analysis to represent how much someone changed or improved over time.

iii. Repeated measures ANCOVA use the baseline value as a covariate in an analysis of covariance (ANCOVA) model.

iv. Repeated measure ANCOVA using baseline as covariate and change scores for repeated measure. This method uses change score model in method 2, adjusted for its baseline.

Results

The descriptive and results from the analyses were presented in Table 1, 2 and 3 below. The multivariate

Table 1: *Descriptive on variable 'b'*

	Experiment n=67		Control n=74		Group n=141			
	Mean	SE	Mean	SE	Mean	SE	95% CI	
Baseline 1	44.26	1.09	43.54	1.17	43.91	0.809	42.3	45.5
Posttest 2	47.72	1.01	44.07	1.14	45.90	0.770	44.3	47.4
Posttest 3	50.35	0.98	44.70	1.23	47.53	0.802	45.9	49.1
	44.26	1.09	43.54	1.17	43.91	0.809	42.3	45.5

Table 2: *Multivariate tests on combined dependent variables*

		Wilks' lambda Value	F	Hypothesis df	Error df	Sig	Partial Eta Squared	Observed Power[a]
Method 1	Group	0.996	0.0049	7	133	0.103	0.08	0.694
ANOVA _raw	Time *Group	0.768	2.714	14	126	0.002	0.23	0.987
Method 2	Group	0.784	5.221	1	139	0.000	0.24	0.997
ANOVA_change	Time *Group	0.957	0.848	1	139	0.550	0.04	0.355
Method 3	*Group*	*0.83*	*3.73*	*7*	*126*	*0.001*	*0.17*	*0.973*
ANCOVA-raw	*Time *Group*	*0.91*	*1.81*	*7*	*126*	*0.090*	*0.09*	*0.710*
Method 4	*Group*	*0.79*	*4.52*	*7*	*126*	*0.000*	*0.20*	*0.990*
ANCOVA_change	*Time *Group*	*0.96*	*0.68*	*7*	*126*	*0.680*	*0.04*	*0.290*

a. Computed using alpha = .05

Table 3: *The univariate test using the measure of focus: variable 'b'*

	Mean Square	F	Sig	Partial Eta Squared	Power
Method 1	1180.20	5.918	0.016	0.04	0.676
Method 2	214.20	9.218	0.003	0.06	0.922
Method 3	596.20	13.702	0.000	0.09	0.957
Method 4	324.39	14.650	0.000	0.10	0.970

statistical tests were run and the key statistic in multivariate is the F-test of difference of group means was reported. F test if the means of the groups (n=141) formed by values of the combined values of the seven multiple independent variables were different enough not to have occurred by chance. Where the group means do not differ significantly then it is inferred that the independent variable (i.e. Intervention) did not have an effect on the dependent variable. Although there were seven dependent variables DVs, only one (i.e. variable b) will be used as the focus of discussion. The significance levels, effect size and power were reported as one should always report effect sizes as well as significance level when reporting ANOVA results (7).

With method 1, the Multivariate between-subject test did not reach statistical significant result on the combined dependent variables, even though the within subject's interaction between time and group reached significance. Close examination of the univariate test show only one variable, b: F(1, 139) =5.9, *p*=0.02, partial eta squared=0.04 and power of 65% reached statistical significance at p<o.o5 level. This variable, b will be used as the focus for comparison on all the method used.

With method 2, the Multivariate between-subject test reach statistical significant result on the combined dependents, with the within subject's interaction between time and group almost reaching significance but was not significant at p<0.05 level. Close examination of the univariate test show six out of seven variables reached statistical significant. The variable of interest, b: F(1, 136) =9.22, *p*=0.003, partial eta squared=0.06 and power of 85% reached statistical significance at p<0.05 level.

With method 3, the Multivariate between-subject test reach statistical significant result on the combined dependents F(7, 126) =3.74,, *p*=0.001, partial eta square=0.17, power of 97 percent, whilst the within subject's interaction between time and group did not reach statistical significance at p<0.05 level. Close examination of the univariate test show six out of seven seven reached statistical significance. The variable b: F(1,132) =13.7, *p*=0.000, partial eta squared=0.09 and power of 96% reached statistical significance at p<0.05 level.

With method 4, the Multivariate between-subject test reach statistical significant result on the combined dependents F(7, 126) =4.52,, *p*=0.000, partial eta square=0.2, power of 99 percent, whilst the within subject's interactional effect between time and group did not reach statistical significance at p<0.05 level. Close examination of the univariate test on the seven dependents also show six out of seven reached statistical significance. The variable b: F(1,126) =14.7, *p*=0.000, partial eta squared=0.10 and power of 97% reached statistical significance at p<0.001 level.

Discussion

The use of ANCOVA was useful for the goal of reducing the error term in the analyses model. I will describe at least three purposes. First, it is used in quasi-experimental designs, to remove the effects of variables which modify the relationship of the categorical independents to the interval dependent. Second, in experimental designs, it is used to control for factors which cannot be randomized but which can be measured on an interval scale. Statistical expert suggest the addition of covariates to a model is rarely needed in experimental research, as randomization in principle controls for all measured and unmeasured confounding variables. Nevertheless, if any covariates are added, they must be correlated with the treatment (independent) variable for uncorrelated covariates complicates interpretation of results as in principle it is controlling for something already controlled for by randomization. However, if the covariate is correlated with the treatment/independent, then its inclusion will lead the researcher to underestimate of the effect size of the treatment factors (independent variables). Third, ANCOVA is also used for regression models, to fit regressions where there are both categorical and interval independents.

With this basic overview of ANCOVA, it is easier to understand why the method 3 and 4 are the most efficient model for analyzing repeated measures in this study. Of the four methods, method I and 2 does not adjust the baseline at all. Method I is the worst method to used and has the highest p value, lowest effect size ands lowest power. The result of the finding is also very different from all the results of the other three methods. Method 2-4 had almost

similar result suggesting they are more accurate and reliable analyses on this data base. Method 2 although it attempt to represent the most accurate change over the two time period, however it still did not consider adjusting the baseline values. Method 3 and 4 uses ANCOVA and appears to be the most reliable methods of analyses. Vickers (8) assert that ANCOVA has the highest statistical power. Of the two, method 3 has the lowest Wilk's Lambda value (9) , which suggests the greatest differences between the groups being studied. However, it seems that method 4 is the best as it consider the adjusted baseline values as well as the correct change in the repeated measure. It has the highest power and effect size.

Conclusions

Reporting the changes from baseline is fundamental. The statistical calculation of significance must be based on adjusted values to take into account the differences between the two groups to be compared. Although the result of a *t-test* to examine the differences between the mean scores of the two groups at baseline were found not statistically significant, i.e. that they were comparable, but this does not mean that the imbalance between the group are negligible. Thus, using Analysis of Covariates (ANCOVA) ensure we correct the imbalance between the group at baseline. Based on the findings, method I should not be used at all. In cases where ANCOVA cannot be used such as in very small sample or where there are violations of the assumptions underlying ANCOVA models, the best method to use would be the change scores. The finding presented here support the use of ANCOVA as the method of choice for analysing the result of trials with baseline and repeated measures that meet the normality assumption.

Acknowledgements

The paper draws from the study on the SAMA project funded by MAKNA and fundamental grant. The author expressed deep gratitude to Karuthan Chinnai for generously sharing his expertise on biostatistics.

References

1. Lieberman MA, Golant M, Giese-Davis J, Winzlenberg A, Benjamin H, *et al*. Electronic support groups for breast carcinoma. *Cancer* 2003; 97 (4):920-925.

2. Schwartz AL, Mori M, Gao R, *et al*. Exercise reduces daily fatigue in women with breast cancer receiving chemotherapy. *Med Sci Sports Exerc* 2001; 33 (5): 718-723.

3. Jenkins V, Shilling V, Deutsch G, Bloomfield D, *et al*. A 3 year prospective study of the effects of adjuvant treatments on cognition in women with early stage breast cancer. *Br J of Cancer* 2006; 94: 828-834.

4. Salinsky M, Storzbach D, Dodrill CB, Binder LM. Test–retest bias, reliability, and regression equations for neuropsychological measures repeated over a 12–16-week period. *J Int Neuropsychological Soc* 2001; 7 :597-605.

5. Portney LG, Watkins MP. Foundation of Clinical Research Application to Practice. 2nd ed. New Jersey: Prentice Hall Health; 2002.

6. Pallant, J. SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows (Version 12-14). Sydney: Allen & Unwin; 2002.

7. APA Publication manual of the American Psychological Association. 4th ed, Washington, DC: American Psychological Association; 1994.

8. Vicker AJ. The use of percentage change from baseline as an outcome in a controlled trail is statistically inefficient: a simulation study. BMC Medical Research Methodology 2002; 1:6 http://www.biomedcentral.com/1471-2288/1/6. Accessed 15 April 2008.

9. Tabachnick BG, Fidell LS. Using Multivariate Statistic. 5th ed. Boston: Allyn & Bacon; 2007.