Original Article

Management of Baseline Measurements in Statistical Analysis of 2×2 Crossover Trials

Hamid Alavi Majd[1], Parisa Naseri[1], Somayeh Momenyan[1] ✉, Saeideh Heidari[2]

[1]Department of Biostatistics, Faculty of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[2]Department of Nursing, Faculty of Nursing and Midwifery, Qom University of Medical Sciences, Qom, Iran

Abstract

Introduction: Crossover designs have applications in a wide range of sciences. The simplest and most common of such designs are the two-period, two-treatment (2×2) crossover. As a consequence, each subject provides a 4×1 vector of responses for data analysis in the following chronological order: baseline (period 1), post-baseline (period 1), baseline (period 2), and post-baseline (period 2).

Methods: We considered three types of analytic approaches for handling the baselines:1) analysis of variance (ANOVA) method which ignores the first or both period baselines or use a change from baseline analysis 2) analysis of covariance (ANCOVA) method which uses an analysis of covariance where linear functions of one or both baselines are employed as either period-specific or period-invariant covariates 3) Joint modeling method that conducts joint modeling of a linear function of the baseline and post-baseline responses with certain mean constraints for the baseline responses. The crossover clinical trial data was analyzed, using the proposed models.

Results: Based on the results on real data among all mentioned models, the first model (direct comparison of post-treatment values) and the second model (post-treatment measurement subtracts corresponding baseline) had the lowest and the highest standard errors, respectively. With respect to Akaike Information Criterion (AIC), the fifth model (comparison of post-treatment values adjusted by all available baseline data) and the eighth model (comparison of post-treatment values adjusted by difference and sum of all available baseline data) had the lowest magnitude, and the ninth model (modeling period baseline jointly with post-treatment values) had the highest AIC for both variables which the values of AIC were 518.1, 520.9 and 1137.8, respectively.

Conclusion: To sum up, it is found that baseline data of crossover trial may be used to improve the efficiency of treatment effect estimation when applied appropriately.

Keywords: Baseline adjustment; Covariate; Crossover trial; Carryover effect; Change scores

1.Introduction

Crossover designs have applications in a wide range of sciences and research areas, such as clinical trials, pharmaceutical studies, psychological experiments, agriculture field trials, and animal feeding experiments [1]. The advantage of such designs is that the subjects become their own controls, thereby reducing the error variance. The simplest and most common of such designs are the two-period, two-treatment (2×2) crossover, with the

treatments generically labeled A and B. In a typical trial, subjects are randomized in a 1:1 ratio to either the AB sequence (receive treatment A in period 1 and treatment B in period 2) or the BA sequence (receive treatment B in period 1 and treatment A in period 2). A 'washout' of suitable length is imposed between the two periods to minimize the risk of a potential carryover effect of the first treatment [2]. For each subject, a continuous response of interest (e.g., systolic blood pressure) is measured before and after a fixed-duration administration of the assigned treatment within each period. Accordingly, each subject provides a 4×1 vector of responses for data analysis in the following chronological order: baseline (period 1), post-baseline (period 1), baseline (period 2), and post-baseline (period 2), where baseline refers to the pre-treatment response within each period. Using these baseline measurements can improve the statistical power of crossover designs [3].

With the increasing popularity of crossover designs in the past three decades, different analytic approaches for modeling baseline data in crossover trials have been suggested in the literature, such as Hills and Armitage [4], Wallenstein [5], Willan and Pater [6], Kenward and Jones [2], and the recent work by Metcalfe [7] and Kenward and Roger [3]. As such, for the 2×2 crossover, four types of analytic approaches for handling the baselines can be considered: ignore the first or both period baselines, use a change from baseline analysis (very common), use an analysis of covariance where linear functions of one or both baselines are employed as either period-specific or period-invariant covariates, or conduct joint modeling of a linear function of the baseline and post-baseline responses with certain mean constraints for the baseline responses. This study reviews all of the models that have been presented in different studies on 2×2 crossover trial until now [2, 3, 5-11]. The performance of statistical methods will

also be investigated in an analysis of real data from a crossover study of treatments for effectiveness of pistachio nut supplementation on *High-density Lipoprotein* (*HDL*) and Low-density Lipoprotein (LDL) measures in patients with type 2 diabetes. In this study, no carryover effect and unstructured variance-covariance matrix (Σ) was assumed.

2. Materials and Methods
2.1 Modeling methods

For each subject, let *Xi* and *Yi* denote the baseline and post treatment response within period i ($i = 1, 2$), respectively. It is assumed that $E(X_1) = \lambda_1$ and $E(X_2) = \lambda_2$ for both the AB and BA sequences, $E(Y_1) = \lambda_1 + \mu_A$ and $E(Y_2) = \lambda_1 + \mu_B$ for the AB sequence, and $E(Y_1) = \lambda_1 + \mu_B$ and $E(Y_2) = \lambda_1 + \mu_A$ for the BA sequence. This common formulation allows for the true periods 1 and 2 baseline means to be different within each sequence, but it assumes that the period-specific baseline means are the same across the two sequences. This is equivalent to assuming that there is no [differential] carryover effect between periods 1 and 2 for the two treatments. In addition, it is assumed that the true variance-covariance matrix of the 4×1 response vector $(X_1, Y_1, X_2, Y_2)^T$ is sequence invariant, with the general (i.e., unstructured) version denoted by

$$\begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 Y_1} & \sigma_{X_1 X_2} & \sigma_{X_1 Y_2} \\ & \sigma_{Y_1}^2 & \sigma_{X_2 Y_1} & \sigma_{Y_1 Y_2} \\ & & \sigma_{X_2}^2 & \sigma_{X_2 Y_2} \\ & & & \sigma_{Y_2}^2 \end{bmatrix} = \begin{bmatrix} V_1 & \rho_{12}\sqrt{V_1 V_2} & \rho_{13}\sqrt{V_1 V_3} & \rho_{14}\sqrt{V_1 V_4} \\ & V_2 & \rho_{23}\sqrt{V_2 V_3} & \rho_{24}\sqrt{V_2 V_4} \\ & & V_3 & \rho_{34}\sqrt{V_3 V_4} \\ & & & V_4 \end{bmatrix}$$

Different studies have showed that the mean of the six within-subject correlations is typically in the 0.6–0.9 range, with individual correlations ranging from 0.5 to 0.95. Different methods exist for estimating treatment`s effect ($\delta = \mu_A - \mu_B$) and for testing $H_0 : \delta = 0$ versus $H_1 : \delta \neq 0$. All

methods yield an unbiased estimator of δ variance of estimates and therefore their precisions are different. Three general modeling methods to estimate treatment effect δ have been described as follows:

Method 1: ANOVA method

Model 1: direct comparison of post-treatment values:

$$\hat{\delta} = 1/2(Y_1 - Y_2)$$

Model 2: direct comparison of change from baseline (post-treatment measurement subtracts

Corresponding baseline):

$$\hat{\delta} = 1/2((Y_1 - X_1) - (Y_2 - X_2))$$

Method 2: ANCOVA method

Model 3: comparison of post-treatment values adjusted by trial baseline:

$$\hat{\delta} = 1/2(Y_1 - Y_2 \mid X_1)$$

Model 4: comparison of post-treatment values adjusted by corresponding period baseline:

$$\hat{\delta} = 1/2(Y_1 \mid X_1 - Y_2 \mid X_2)$$

Model 5: comparison of post-treatment values adjusted by all available baseline data (across different periods) simultaneously:

$$\hat{\delta} = 1/2(Y_1 - Y_2 \mid X_1 X_2)$$

Model 6: comparison of post-treatment values adjusted by sum of all available baseline data (across different periods)

$$\hat{\delta} = 1/2(Y_1 - Y_2 \mid X_1 + X_2)$$

Model 7: comparison of post-treatment values adjusted by difference of all available baseline data (across different periods)

$$\hat{\delta} = 1/2(Y_1 - Y_2 \mid X_1 - X_2)$$

Model 8: comparison of post-treatment values adjusted by difference and sum of all available baseline data (across different periods) simultaneously:

$$\hat{\delta} = 1/2(Y_1 - Y_2 \mid (X_1 + X_2, X_1 - X_2))$$

Method 3: Joint modeling method:

Model 9: modeling period baseline (X_1, X_2) jointly with post-treatment values (Y_1, Y_2)

Model 10: modeling difference period baseline $(X_1 - X_2)$ jointly with difference post-treatment values $(Y_1 - Y_2)$

The response vector $(X_1, Y_1, X_2, Y_2)^T$ is assumed to follow a four-vitiate normal distribution with $E(X_1) = \lambda_1$ and $E(X_2) = \lambda_2$ for both the AB and BA sequences. Subject to these mean constraints, the means of Y_1 and Y_2 can be estimated within each sequence via a standard restricted maximum likelihood (REML) approach. All statistical analyses were performed in SAS software version 9.2(SAS Institute, Cary, NC). For more accurate comparisons, the results were reported with three decimal places.

2.2 Clinical data

Study was conducted to determine the effect of pistachio nut supplementation on High-density lipoprotein (HDL) and Low-density lipoprotein (LDL) measures in patients with type 2 diabetes [12]. A double blinded, randomized, crossover clinical trial was carried out in Shahid Beheshti Hospital of Qom, Iran, in the period between February 2012 and March 2013. The study protocol for this data was approved by Ethics Committee of Qom University of Medical Sciences. Forty-four patients with type 2 diabetes met the inclusion criteria and were enrolled in the study. 44 diabetic patients were equally assigned to groups A and B. Patients in group A received a snack of 25 g pistachio nuts twice a day for 12 weeks and group B received a control meal without nuts. After 12 weeks of intervention, the patients had an 8-week washout. Then the groups were displaced, and group B received the same amount of pistachios for 12 weeks. At the beginning and end of the first four weeks and second

four weeks, High-density lipoprotein (HDL) and Low-density lipoprotein (LDL) were measured (Figure 1). Carryover effect was

not significant for any of the two variables mentioned above.

3. Results

Measurements of the two groups of diabetic patients in the first and second phase are shown in Table 1. The estimated correlations between vectors of HDL responses $(X_1, Y_1, X_2, Y_2)^T$ were range from 0.74 to 0.83, with a mean of 0.77. Also the estimated correlations between vectors of LDL responses were range from 0.26 to 0.85, with a mean of 0.56 that was lower than HDL.

Summary statistics and analysis details of all models are shown in Table 2 and 3 for HDL and LDL variables respectively.
Among all models mentioned above, the first and the second models had lowest and highest standard errors, respectively for both variables. With respect to Akaike Information Criterion (AIC), fifth and eighth models had the lowest magnitude, and the ninth model had the highest AIC for both variables (Table 2, 3).
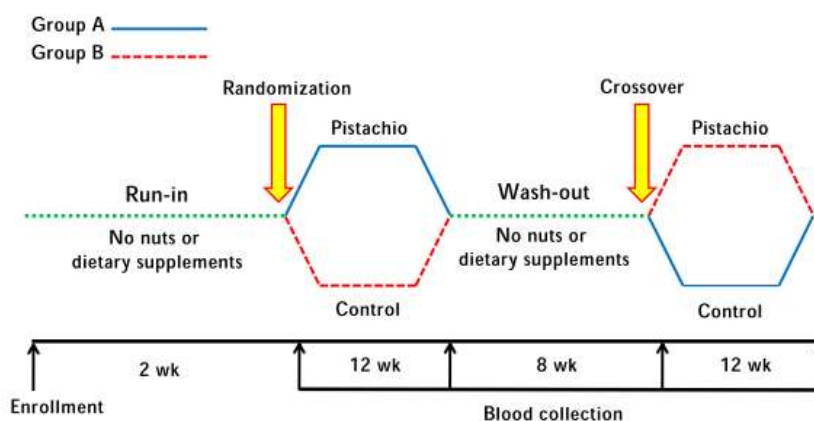

Figure 1. Flow chart of study protocol

Table1. Measurements of the two groups of type 2 diabetic patients in the first and second phase

Parameters	first phase					
	Group B control (n=21)			Group A pistachio (n=23)		
	Before	After	Differences	Before	After	Differences
HDL	58.8 ± 9.8	50.5 ± 8.5	8.3 ± 6.7	58.3 ± 10.1	48.93 ± 9.3	9.4 ± 6.6
LDL	79.1 ± 28.4	88.7 ± 21.4	-9.6 ± 27.1	72.7 ± 25.1	78.8 ± 29.9	-6.3 ± 18.5
	Second phase					
	Group B pistachio (n=21)			Group A control (n=23)		
	Before	After	Differences	Before	After	Differences
HDL	47.1 ± 8.9	54.3 ± 9.6	-7.2 ± 3.7	53.8 ± 9.7	54.4 ± 10.5	-0.6 ± 5.4
LDL	88.7 ± 25.8	86.1 ± 29.9	2.5 ± 19.9	76.5 ± 21.6	75.3 ± 23.7	-0.8 ± 16.3

By considering significant level 0.05, effect of pistachio was significant in second and fourth models while there were no significant effects on other models for HDL variable (Table 2). There were no significant effects of Pistachio on any models for LDL variable but there were considerable differences between them (Table 3). Moreover, in the fifth and eighth model, estimation of the effect of pistachio and its standard error were almost identical. All models showed almost the same estimation of effect of pistachio for both variables (Table 2, 3).

Table 2. summary statistics and analysis details for High-density lipoprotein (HDL)

Type of model	$\hat{\delta}$	$\sqrt{\hat{V}(\hat{\delta})}$	Two-tailed p-value	AIC
First Model	-1.522	0.790	0.060	581.5
Second Model	-2.954	1.240	0.021	548.9
Third Model	-1.557	0.791	0.055	544.7
Fourth Model	-2.655	1.090	0.019	552.3
Fifth Model	-1.237	1.013	0.229	518.1
Sixth Model	-1.344	0.804	0.102	521.8
Seventh Model	-1.510	0.986	0.133	584.9
Eighth Model	-1.237	1.013	0.229	520.9
Ninth Model	-1.237	0.824	0.140	1137.8
Tenth Model	-1.510	0.817	0.071	572.8

Table 3. summary statistics and analysis details for Low-density lipoprotein (LDL)

Type of model	$\hat{\delta}$	$\sqrt{\hat{V}(\hat{\delta})}$	Two-tailed p-value	AIC
First Model	-17.545	17.528	0.322	921.1
Second Model	-20.238	18.322	0.275	922
Third Model	-18.910	17.839	0.295	908
Fourth Model	-6.510	15.892	0.275	917.4
Fifth Model	-16.363	18.312	0.377	870.6
Sixth Model	-18.198	18.149	0.312	886.3
Seventh Model	-15.661	17.703	0.381	915.4
Eighth Model	-16.363	18.312	0.377	873.4
Ninth Model	-16.363	18.169	0.372	1666.3
Tenth Model	-15.661	17.768	0.382	924.4

4. Discussion

Crossover designs are typically used as relative to parallel group designs; statistical efficiency can be gained through leveraging the within-subject correlations among responses to different treatments. To maximize such gains when baseline (pre-treatment) responses are collected within each period of a crossover, it is imperative that careful consideration be given to how those data will be used in the analysis. The present study has discussed many different models to tackle these baseline values through focusing on the simple 2×2 crossover. The real data was the clinical study that surveyeed effect of pistachio nut supplementation on High-density lipoprotein (HDL) and Low-density lipoprotein (LDL) measures in patients with type 2 diabetes. Diabetes mellitus is a chronic disease that affects about 5-10% of the world population [13]. Diet and weight control are the basic principles in the treatment of type 2 diabetes. Nuts contain magnesium and monounsaturated and polyunsaturated fats, which are supposed to improve insulin sensitivity, carbohydrate metabolism, and insulin homeostasis [14]. With respect to the results on standard error of effect of pistachio diet on HDL and LDL

parameters, a common approach (method II) which uses the change values in each period is an inappropriate model among of all models.

Mehrotra in his study compared variance of different models in different structure of the variance-covariance matrix of data. He showed that the second model will have a larger variance than other models theoretically [9].

Chen et al. in their study used simulation and manifested that the estimated standard error of the second model is the largest under unstructured variance-covariance matrix. They also indicated that the first and third models have smallest standard error among others [8] whereas in the present study, the first and the fourth model had smallest standard error.

Mehrotra also illustrated that under compound symmetry structure assumption, the first model and under other structures, seventh model have lowest variance, theoretically [9].

The present study results showed the ninth model has largest AIC among all models. Also, in Mehrotra study, using the ninth model was not recommended because of type I error rate. He similarly showed that using simulation in the tenth and seventh models have best Type I error rate and power. Moreover, seventh model delivers better control of type I error rate than tenth model in small samples [9].

Based on between and within period correlations, only for LDL, within period correlations were upper than between period correlations while HDL factor had high between and within period correlations.

Chen et al. in their study showed when data are strongly correlated within the same period but weakly correlated in different periods, the efficiency of all models was the high but when between period correlations were high, the second and third models had low efficiency and joint modeling methods had higher efficiency to other models. In other words, a unique model can't be

chosen as the best model and based on different structures of the variance-covariance matrix, various models will have different efficiency [8].

In this study, only 2×2 crossover design was investigated for a real data. It is highly recommended that all the methods proposed for crossover design compare respect to type I error, power, AIC and standard errors using simulation. Although our focus has been on the 2×2 crossover design, all the competing methods can be easily extended to higher-order crossover designs and according to the different criteria, comparison can be conducted among them.

5. Conclusion

There is no obviously 'best' method to tackle baselines for such as a two-period crossover design and based on different structures of the variance-covariance matrix, various models will have different efficiency.

Conflict of Interest
The authors declare no conflict of interest.

References
1.Kenward MG. Crossover Design. Wiley Encyclopedia of Clinical Trials.
2.Kenward MG, Jones B. The analysis of data from 2× 2 cross-over trials with baseline measurements. Statistics in Medicine. 1987;6(8):911-26.
3.Kenward MG, Roger JH. The use of baseline covariates in crossover studies. Biostatistics. 2009:kxp046.
4.Hills M, Armitage P. The two-period cross-over clinical trial. British journal of clinical pharmacology. 2004;58(7):S703-S16.
5.Wallenstein S, editor Analysis of the crossover design with baseline. American Statistical Association, Proceedings of the Biopharmaceutical Section; 1986.
6.Willian AR, Pater JL. Using baseline measurements in the two-period crossover clinical trial. Controlled clinical trials. 1986;7(4):282-9.

Journal of Paramdedical Sciences is an open access article under the terms of the Creative Commons Attribution -NonCommercial 4.0 International License

18

7.Metcalfe C. The analysis of cross-over trials with baseline measurements. Statistics in medicine. 2010;29(30):3211-8.

8.Chen X, Meng Z, Zhang J. Handling of baseline measurements in the analysis of crossover trials. Statistics in medicine. 2012;31(17):1791-803.

9.Mehrotra DV. A recommended analysis for 2× 2 crossover trials with baseline measurements. Pharmaceutical statistics. 2014;13(6):376-87.

10.Patel H. Use of baseline measurements in the two-period crossover design. Communications in Statistics-Theory and Methods. 1983;12(23):2693-712.

11.Yan Z. The impact of baseline covariates on the efficiency of statistical analyses of crossover designs. Statistics in medicine. 2013;32(6):956-63.

12.Parham M, Heidari S, Khorramirad A, Hozoori M, Hosseinzadeh F, Bakhtyari L, et al. Effects of pistachio nut supplementation on blood glucose in patients with type 2 diabetes: a randomized crossover trial. Rev Diabet Stud. 2014;11(2):190-6.

13.Jiang R, Manson JE, Stampfer MJ, Liu S, Willett WC, Hu FB. Nut and peanut butter consumption and risk of type 2 diabetes in women. JAMA 2002. 288(20):2554-2560.

14.Lovejoy JC, Most MM, Lefevre M, Greenway FL, Rood JC. Effect of diets enriched in almonds on insulin action and serum lipids in adults with normal glucose tolerance or type 2 diabetes. Am J Clin Nutr 2002. 76(5):1000-1006.