Contemporary Clinical Trials Communications 11 (2018) 20–29

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Comparison of intent-to-treat analysis strategies for pre-post studies with loss to follow-up

Wenna Xi[a], Michael L. Pennell[a,*], Rebecca R. Andridge[a], Electra D. Paskett[b]

[a] Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH 43210, United States
[b] Department of Internal Medicine, College of Public Health, Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, United States

ARTICLE INFO

Keywords:
ANCOVA
Missing data
Mixed model
Multiple imputation
Randomized trial

ABSTRACT

In pre-post studies when all outcomes are completely observed, previous studies have shown that analysis of covariance (ANCOVA) is more powerful than a change-score analysis in testing the treatment effect. However, there have been few studies comparing power under missing post-test values. This paper was motivated by the Behavior and Exercise for Physical Health Intervention (BePHIT) Study, a pre-post study designed to compare two interventions on postmenopausal women's walk time. The goal of this study was to compare the power of two methods which adhere to the intent-to-treat (ITT) principle when post-test data are missing: ANCOVA after multiple imputation (MI) and the mixed model applied to all-available data (AA). We also compared the two ITT analysis strategies to two methods which do not adhere to ITT principles: complete-case (CC) ANCOVA and the CC mixed model. Comparisons were made through analyses of the BePHIT data and simulation studies conducted under various sample sizes, missingness rates, and missingness scenarios. In the analysis of the BePHIT data, ANCOVA after MI had the smallest *p*-value for the test of the treatment effect of the four methods. Simulation results demonstrated that the AA mixed model was usually more powerful than ANCOVA after MI. The power of ANCOVA after MI dropped the fastest as the missingness rate increased; in most simulated scenarios, ANCOVA after MI had the smallest power when 50% of the post-test outcomes were missing.

1. Introduction

In a pre-post study a treatment is evaluated by measuring responses both before and after the study for each participant in a treatment group and a control group. Pre-post study designs have been widely used in clinical trials, psychology, education, and sociology. For example, our research was motivated by the Behavior and Exercise for Physical Health Intervention (BePHIT) Study, a pre-post study designed to compare two interventions intended to promote walking in postmenopausal women [1].

When there is complete follow-up, previous studies have shown that, in terms of testing the treatment effect, analysis of covariance (ANCOVA) is more powerful than a comparison of change scores [2–5]. However, in reality, missing data, in particular loss to follow-up, is very common in pre-post studies. For instance, in the BePHIT study, 17% of the participants did not finish the study. With unbalanced sample sizes for pre- and post-test levels in each treatment group, a regular ANCOVA or change score analysis cannot be conducted without dropping any subjects. Therefore the most straightforward method for handling missing values is to exclude all the subjects with missing data. This type of analysis is called the complete-case (CC) analysis. The CC analysis is usually not recommended, since it throws away information collected in the study and does not follow the intent-to-treat (ITT) principle for clinical trials [6,7]. Nowadays, one popular way to handle missing data is multiple imputation (MI) [6]. For instance, in pre-post studies, missing follow-ups can be simulated multiple times using the baseline outcome value and measured covariates and the results of the analysis of each complete data set are combined to account for the uncertainty introduced by the imputations [8]. Another approach often used for data with repeated measures is the mixed model, where all available pre- and post-test values are regressed over treatment and timepoint indicators, assuming some variance-covariance structure for the repeated measures.

The main goal of this study was to compare the power of two analysis methods which adhere to ITT principles: the mixed model and ANCOVA after MI for pre-post studies when missing post-test is present. We also wanted to compare these methods to two methods that do not adhere to an ITT principle: ANCOVA and the mixed model using only completely observed cases. These methods were first compared in the context of our motivating example (BePHIT) and then in simulation

* Corresponding author.
E-mail address: pennell.28@osu.edu (M.L. Pennell).

https://doi.org/10.1016/j.conctc.2018.05.008
Received 29 November 2017; Received in revised form 11 April 2018; Accepted 3 May 2018
Available online 09 May 2018

studies based on the BePHIT data. A parallel set of simulations comparing the type I error rate of the four methods were conducted as well.

2. Motivating example

The Behavior and Exercise for Physical Health Intervention (BePHIT) Study was a randomized controlled study of a 12-week walking intervention conducted on postmenopausal women between January 2008 and March 2009 [1]. The primary outcome was the change in time for women to finish a one-mile walk. In addition to one-mile walk time, anthropometric and psychometric measures were obtained at pre- and post-test.

After passing the selection criteria, 71 participants were stratified by BMI and randomized into either a coach group or a no-coach group. For women in the coach group, a trained coach was assigned. The role of the coach was to explain the intervention, provide the first week's steps goal, train subjects to use a pedometer and the Interactive Voice Response (IVR) system to collect data, and offer help during the intervention. Women in the no-coach group received similar instructions, training, and help, except that they were not informed that they had access to a coach. Although both groups received a treatment, to be consistent with the terminology in this paper, we will refer to the coach condition as the treatment and the no coach condition as the control.

Among the 71 randomized participants, 35 were assigned to the treatment group and 36 to the control group. For the control group, baseline walking time was only available for 35 patients. In total, 12 (17%) patients dropped out before the post walking test, 4 of whom were in the treatment group and 8 in the control group. The drop out rate did not differ significantly across the two groups ($p = 0.20$). The original study reported 2 withdrew and 12 did not complete in the treatment group, and 7 withdrew and 11 did not complete in the control group [1]. In that study, “completed” was defined as completing all post-test assessments within 30 days after the end of the walking intervention. These post-test assessments included the walk test and anthropometric and psychometric measures not considered in this paper. Those who had their post-test score recorded but did not finish all their anthropometric and psychometric measures were also included in this analysis, which led to fewer dropping out here.

3. Analysis methods for pre-post studies with complete data and missing data

3.1. Pre-post studies

A pre-post study is a randomized controlled study where outcome values are measured both before and after the study. As opposed to treatment-control studies where the outcome variable is only measured once, pre-post studies allow investigators to account for the level of the outcome variable before the treatment is applied. Different from a one-group pre-post design, a treatment-control pre-post study controls for secular trends [9,10]. In the BePHIT study, for instance, besides the intervention, the improvement of women's one-mile walk time may have been caused by some other factors, such as a national walking campaign, affecting the women during the same time period. A one-group pre-post study fails to consider these factors; however a treatment-control pre-post study accounts for secular trends by comparing the results from the treatment group to a control group observed over the same period of time.

If we let $\mathfrak{R}$ be the randomization process, $\mathfrak{T}$ be the treatment process, and ($Y_{pre,t}$, $Y_{post,t}$) and ($Y_{pre,c}$, $Y_{post,c}$) be the pre- and post-test measure of a treated and control participant, respectively, then a pre-post study design can be illustrated by the following:

$$\begin{aligned} \mathfrak{R} &\rightarrow Y_{pre,t} \overset{\mathfrak{T}}{\rightarrow} Y_{post,t} \\ \mathfrak{R} &\rightarrow Y_{pre,c} \rightarrow Y_{post,c} \end{aligned}$$

3.2. Analysis of pre-post studies with complete data

Many analysis approaches for pre-post studies have been discussed [2,3,5,9,11,12]. Arguably the two most common analysis methods are the change score analysis and Analysis of Covariance (ANCOVA) [2]. We discuss these two methods and their statistical power in the following.

3.2.1. Change score analysis

A change score analysis first obtains the difference in outcome values before and after the experiment, and then regresses the difference on the treatment assignment using the following model:

$$Y_i - X_i = \alpha_{C0} + \alpha_{C1} T_i + \varepsilon_i, \tag{1}$$

where Y_i is the post-test outcome level for subject i, X_i is the pre-test outcome level for subject i, T_i is the indicator variable for treatment assignment, and ε_i is the error term for subject i ($\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$). Note that $Y_i - X_i$ is the change score for subject i during the experiment. In the model above, α_{C1} quantifies the effect of treatment assignment on change in outcome level from pre to post. Since T_i is a binary variable, the change score test is equivalent to a two-sample t-test comparing the mean of $Y_i - X_i$ between treatment and control groups.

3.2.2. ANCOVA

Unlike the change score method, in ANCOVA, post-test value (Y_i) is treated as the outcome variable and pre-test value (X_i) is treated as a predictor. The ANCOVA model can be expressed as:

$$Y_i = \alpha_{A0} + \alpha_{A1} T_i + \alpha_{A2} X_i + \varepsilon_i, \tag{2}$$

where T_i and ε_i are as defined in the change-score model. ANCOVA assumes that pre-test values are measured without error [5]. This assumption holds for variables, such as weight and height, that can be measured precisely. However, it is often violated for self-reported measurements and educational or psychological tests. In the model above, α_{A1} is the effect of treatment assignment on the post-test scores adjusting for the pre-test scores.

The ANCOVA model (2) can also be written as

$$Y_i - X_i = \alpha_{A0} + \alpha_{A1} T_i + \alpha_{A2}^* X_i + \varepsilon_i, \tag{3}$$

where $\alpha_{A2}^* = \alpha_{A2} - 1$ from (2) [2,5]. Thus, ANCOVA can be viewed as an extension of the change score model (1) to include pre-test level X_i as a predictor.

3.2.3. Power comparison: change score analysis vs. ANCOVA

Oakes and Feldman compared the detectable treatment effects of the change score and ANCOVA models [5]. Assume

1. $Var(X_i) = Var(Y_i) = \sigma^2$ regardless of the experimental group,
2. Number of subjects in each group is the same, and
3. Pre-test is measured without error.

Under the assumption of normally distributed errors, the detectable treatment effect for the change score analysis at type-I and type-II error rates of α and β is

$$\Delta_C = \sqrt{\frac{4\sigma^2(1-\rho)(Z_{1-\alpha/2} + Z_{1-\beta})^2}{m}},$$

and the detectable treatment effect for ANCOVA is

$$\Delta_A = \sqrt{\frac{2\sigma^2(1-\rho^2)(Z_{1-\alpha/2} + Z_{1-\beta})^2}{m}},$$

where $\rho = Corr(X_i, Y_i)$, Z_x is the xth quantile of the standard normal distribution, and m is the number of subjects in each experimental group. Therefore we have