Contemporary Clinical Trials 37 (2014) 225–233

Contents lists available at ScienceDirect

Contemporary Clinical Trials

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

A comparison of nonparametric and parametric methods to adjust for baseline measures

Martin O. Carlsson [a,b,*], Kelly H. Zou [a], Ching-Ray Yu [a], Kezhen Liu [b], Franklin W. Sun [a]

[a] Pfizer Inc., New York, NY, USA
[b] Department of Statistics, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA

ARTICLE INFO

Article history:
Received 13 July 2013
Received in revised form 8 January 2014
Accepted 13 January 2014
Available online 21 January 2014

Keywords:
Randomized controlled trials
Percent change from baseline
Analysis of covariance
Robust regression
Covariate imbalance

ABSTRACT

When analyzing the randomized controlled trial, we may employ various statistical methods to adjust for baseline measures. Depending on the method chosen to adjust for baseline measures, inferential results can vary. We investigate the Type 1 error and statistical power of tests comparing treatment outcomes based on parametric and nonparametric methods. We also explore the increasing levels of correlation between baseline and changes from the baseline, with or without underlying normality. These methods are illustrated and compared via simulations.

1. Introduction

For randomized controlled trials, inferential results may vary depending on whether or not an individual subject's baseline data are adjusted for, as well as the method chosen to adjust. Either the post-treatment value or change from baseline may be analyzed to account for the effect of the baseline measures. Alternatively, the percent change from baseline, a scale invariant method can also be used. It is a measure that is easy to interpret although its distribution is complicated, especially when baseline and post-baseline measures are correlated.

Lord's paradox states that the relationship between a continuous outcome and a categorical exposure may be reversed when an additional continuous covariate (e.g., baseline measures) is introduced [1]. Thus, appropriate inferential procedures must be employed to adjust for baseline measures. The analysis of covariance (ANCOVA) approach remains the most popular tool in practice even with a set of stringent assumptions including linearity, parallelism, homoscedasticity, and normality.

Previously, Vickers [2] has compared several parametric *t*-test and ANCOVA based methods in a large sample setting. He recommended the use of ANCOVA which had the greatest power in the case of normally-distributed data, when baseline and post-baseline data are correlated. We aim to extend his comparisons by incorporating the strength of correlation and distributional assumptions.

The remainder of this manuscript is organized as follows. In Section 2, we review commonly-used nonparametric methods and parametric univariate methods to compare change scores and percent change from baseline. We also compare different multivariate regression methods to adjust for the baseline. In Section 3, we present Monte-Carlo simulation studies that investigated Type 1 error and statistical power performance. In Section 4, we apply these methods to two published examples containing laboratory assay data and dental caries data. Finally,

* Corresponding author at: 235 East 42nd Street, Mail Stop: 219/7/1, New York, NY 10017, USA. Tel.: +1 212 733 0087; fax: +1 212 309 4368.

E-mail addresses: Martin.Carlsson@pfizer.com (M.O. Carlsson), Kelly.Zou@pfizer.com (K.H. Zou), Ching-Ray.Yu@pfizer.com (C.-R. Yu), LiuKeZhen@gmail.com (K. Liu), Franklin.Sun@pfizer.com (F.W. Sun).

1551-7144/$ – see front matter
http://dx.doi.org/10.1016/j.cct.2014.01.002

Section 5 presents conclusions and discussion based on this research.

2. Univariate and multivariate adjustment methods

2.1. Notations

Let the i-th bivariate measurements under each of treatment groups $k(k = 1,...,K)$ have a bivariate joint distribution function, $H_k(,)$.

$$(X_{ki}, Y_{ki}) \sim i.i.d.H_k(x, y), i = 1, \ldots, n_k, \quad (1)$$

with baseline variable X and post-baseline variable Y, respectively. The total sample size across all k treatment groups is $N = \sum_{i=1}^{K} n_k$.

The post-baseline scores are denoted by Y_{ki}. The change from the baseline is $D_{ki} = Y_{ki} - X_{ki}$. The percent change from the baseline, $(D_{ki}/X_{ki}) \times 100\%$ [3], is recommended for label claim purposes based on patient-reported outcomes [4]. The latter may be easier to interpret since the percent change score is a dimensionless measure.

2.2. Univariate methods

The commonly-used univariate nonparametric method is either the Wilcoxon's rank-sum test (for two samples) or Kruskal–Wallis test (for more than two samples). Alternatively, a parametric method, the two-sample t-test or the one-way Analysis of Variance (ANOVA) may be conducted. It is assumed here that the reader is familiar with the methods being investigated and thus they are not described in detail.

2.3. Multivariate methods

2.3.1. Parametric ANCOVA

The ANCOVA is a commonly-used multivariable regression method on a set of baseline covariates, which may also be conducted to adjust the baseline values. There are five assumptions for conducting the ANCOVA analysis, by expanding on those listed in Rutherford [5]: (1) normality of residuals; (2) homogeneity of variances; (3) homogeneity of regression slopes; (4) linearity of dependent and covariates; (5) independence of error terms.

It is debatable how much deviation from normality of the residuals permitted for ANCOVA to work well. For example, Rutherford [5] and Wilcox [6] have noted that departure from the normality can impact statistical power. In the literature, however, ANOVA, as well as ANCOVA, is robust despite such a departure. Recent discussions on alternative modern robust methods can be found in Erceg-Hurn [7] concerning the consequences due to the departures from normality.

2.3.2. Nonparametric Quade's ANCOVA method

Alternatively, Quade's ANCOVA is conducted by ranking the data marginally across all treatment groups while disregarding the treatment assignments [8–10]. Marginal ranks are denoted as $R(X)$ and $R(Y)$, each by pooling across all treatment groups k. These ranks take on values from 1 to N. In case of ties, the average ranks between the ties are used. These two sets of rankings are then corrected by subtracting the expected rank, $E(R) = (N + 1) / 2$, from each set of rankings.

A linear regression of the adjusted ranks of $\{R(Y) - R(R)\}$ is performed for all data to obtain the residuals. A one-way ANOVA is then performed on these residuals. The p-value is based on an F-distribution with $(k - 1)$ and $(N - k)$ degrees of freedom.

2.3.3. Nonparametric robust regression method

For non-normal data, either the Huber type of M-estimator via the R (http://www.r-project.org) package 'MASS' or Cauchy types in SAS (http://www.sas.com) via Proc Robustreg may be adopted [11–13].

The M-estimator minimizes the sum of a function of the residuals r_i of Y_i rather than the residual squares as in the least-squares method. The effect due to outliers is reduced once we replace the squared residuals $r_i^2$with the following objective function for minimization:

$$min\sum_{i=1}^{N}\rho(r_i), \quad (2)$$

where ρ is a symmetric, positive-definite function with a unique minimum at zero, and is chosen to be non-increasing.

The first derivative is the influence function, i.e.,

$$\psi(y) = \frac{d\rho(y)}{dy}. \quad (3)$$

The weight function is given by:

$$w(y) = \frac{(y)}{y}. \quad (4)$$

The system of equations may be solved if the following iterated reweighted least-squares method is used:

$$min\sum_{i=1}^{N} w\left(r_i^{(l-1)}\right) r_i^2, \quad (5)$$

where $(l - 1)$ indicates the previous iteration before l. The weight $w(r_i^{(l-1)})$ is computed during each of the iterations before the next iteration until convergence.

For the M-estimator of the Huber type,

$$\rho(y) = \begin{cases} \frac{y^2}{2}, \text{if } |y| \le c, \\ c\left(|y| - \frac{c}{2}\right), \text{otherwise.} \end{cases} \quad (6)$$

$$\psi(y) = \begin{cases} y, \text{if } |y| \le c, \\ c \cdot \text{sgn}(y), \text{otherwise.} \end{cases} \quad (7)$$

$$w(y) = \begin{cases} 1, \text{if } |y| \le c, \\ \frac{c}{|y|}, \text{otherwise.} \end{cases} \quad (8)$$