Different ways to estimate treatment effects in randomised controlled trials

Twisk J[a,*], Bosman L[a], Hoekstra T[a,b], Rijnhart J[a], Welten M[a], Heymans M[a]

[a] Department of Epidemiology and Biostatistics, VU Medical Centre, Amsterdam, the Netherlands
[b] Department of Health Science, Faculty of Earth and Life Sciences, VU University, Amsterdam, The Netherlands

ABSTRACT

Background: Regarding the analysis of RCT data there is a debate going on whether an adjustment for the baseline value of the outcome variable should be made. When an adjustment is made, there is a lot of misunderstanding regarding the way this should be done. Therefore, the aims of this educational paper are: 1) to explain different methods used to estimate treatment effects in RCTs, 2) to illustrate the different methods with a real life example and 3) to give an advise on how to analyse RCT data.
Methods: Longitudinal analysis of covariance, repeated measures analysis in which also the baseline value is used as outcome and the analysis of changes were theoretically explained and applied to an example dataset investigating a systolic blood pressure lowering treatment.
Results: It was shown that differences at baseline should be taken into account and that regular repeated measures analysis and regular analysis of changes did not adjust for the baseline differences between the groups and therefore lead to biased estimates of the treatment effect. In the real life example, due to the differences at baseline between the treatment and control group, the different methods lead to different estimates of the treatment effect.
Conclusion: Regarding the analysis of RCT data, it is advised to use longitudinal analysis of covariance or a repeated measures analysis without the treatment variable, but with the interaction between treatment and time in the model.

1. Introduction

Within epidemiology a randomised controlled trial (RCT) is considered to be the best way to investigate the effect of a new treatment. Regarding the analysis of RCT data there is a debate in the epidemiological and biostatistical literature, whether an adjustment for the baseline value of the outcome variable should be made [1–6]. Researchers against this adjustment argue that all differences at baseline between the two groups are due to chance and an adjustment for chance is not correct. Researchers in favour of the adjustment argue that an adjustment is necessary to take into account regression to the mean [7–10]. When differences at baseline between the treatment and control group are due to random fluctuations and measurement error, there is a tendency of the average value to go down in the group with the initial highest average value and to go up in the group with the initial lowest average value. This tendency is known as regression to the mean. Suppose that we are performing an intervention study aiming to improve physical activity among children, and that the intervention has no effect at all. Suppose further that at baseline the intervention group has a lower average physical activity level compared to the control

group. When no adjustment is made for the baseline differences in the outcome variable, in this particular situation, an artificial intervention effect will be estimated. Due to regression to the mean, the average value of the intervention group tend to increase, while the average value of the control group tend to decrease, leading to this artificial intervention effect. When the control group has the higher average value at baseline, the exact opposite occurs: if there is an actual treatment effect in this situation, it will be underestimated due to regression to the mean. In an RCT, regression to the mean can play a major (confounding) role, because the two groups are randomised from one source population. The consequence of this is that they are expected to have the same average baseline value, i.e. the differences between the two groups at baseline are completely due to random fluctuations and measurement error.

Although it seems that the debate is ended in favour of an adjustment for baseline value of the outcome variable, in the literature there are still many RCT's that do not adjust for the baseline values of the outcome variable [11]. Moreover, in the CONSORT statement, which provides guidelines for reporting results of RCTs, there is no statement about the preferred way of analysing RCT data and whether or not an

Table 1
Data structure needed to perform a longitudinal analysis of covariance.

id	Outcome	time	Treatment (*X*)	Baseline
1	Y_{t1}	0	1	Y_{t0}
1	Y_{t2}	1	1	Y_{t0}

adjustment for the baseline value should be made.

When an adjustment is made for the baseline value of the outcome variable, there is a lot of misunderstanding regarding the best way of performing this adjustment. Therefore, the aims of the present educational paper are: 1) to explain different methods used to estimate treatment effects in RCTs, 2) to illustrate the different methods with a real life example and 3) to give an advise on how to analyse RCT data.

2. Methods

2.1. Different methods

The following three statistical methods are mostly used to estimate treatment effects in RCTs: longitudinal analysis of covariance (method 1), repeated measures analysis (method 2) and the analysis of changes (method 3). In the explanation of the different methods, two follow-up measurements are considered. However, the methods can be easily extended with more follow-up measurements.

2.1.1. Method 1: Longitudinal analysis of covariance

Table 1 shows the structure of the data used to estimate the parameters for a longitudinal analysis of covariance.

In this method the outcome variable measured at the different follow-up measurements is adjusted for the baseline value of the outcome (equation (1a)).

$$Y_t = \beta_0 + \beta_1 X + \beta_2 Y_{t0} \tag{1a}$$

where, Y_t = the outcome measured at the two follow-up measurements, X = treatment variable, β_1 = overall treatment effect and Y_{t0} = outcome variable measured at baseline.

To assess the effect of the treatment at the different follow-up measurements, time and the interaction between the treatment variable and time are added to the model (equation (1b)).

$$Y_t = \beta_0 + \beta_1 X + \beta_2 Y_{t0} + \beta_3 time + \beta_4 X \times time \tag{1b}$$

In this model, the regression coefficient for the treatment variable reflects the treatment effect at the first follow-up measurement. The treatment effect at the second follow-up measurement is calculated as the sum of the regression coefficient for the treatment variable and the regression coefficient for the interaction between the treatment variable and time ($\beta_1 + \beta_4$).

2.1.2. Method 2: Repeated measures

Table 2 shows the structure of the data used to estimate the parameters of a repeated measures analysis.

In the repeated measures analysis, the values of all three measurements of the outcome variable (i.e. the baseline value as well as the two follow-up measurements) are used as outcome in the analysis. The model includes time, which is either continuous when a linear

Table 2
Data structure needed to perform the analyses described in method 2.

Id	outcome	time	treatment	baseline
1	Y_{t0}	0	1	Na
1	Y_{t1}	1	1	Na
1	Y_{t2}	2	1	Na

Na = not applicable.

development over time is assumed or represented by dummy variables when a non-linear development over time is assumed (because all three measurements are used as outcome, two dummy variables are needed to represent time) and the interaction between treatment and time (equations (2a) and (2b)).

$$Y_t = \beta_0 + \beta_1 X + \beta_2 time + \beta_3 time \times X \tag{2a}$$

$$Y_t = \beta_0 + \beta_1 X + \beta_2 dummytime_1 + \beta_3 dummytime_2 + \beta_4 dummytime_1 \times X + \beta_5 dummytime_2 \times X \tag{2b}$$

In model 2a, the regression coefficient for the treatment variable reflects the differences between the groups at baseline. To obtain the overall treatment effect over time, time must be coded 1 for both follow-up measurements. The sum of the regression coefficient for the treatment variable and the regression coefficient for the interaction between the treatment variable and time then reflects the overall treatment effect. In the model with the two dummy variables (equation (2b)), the treatment effect at the first follow-up measurement is calculated as the sum of the regression coefficient for the treatment variable and the regression coefficient for the interaction between the treatment variable and the first dummy variable for time ($\beta_1 + \beta_4$), while the treatment effect at the second follow-up measurement is calculated as the sum of the regression coefficient for the treatment variable and the regression coefficient for the interaction between the treatment variable and the second dummy variable for time ($\beta_1 + \beta_5$).

An assumed advantage of repeated measures analysis is that subjects with only a baseline value, but with missing data at all the follow-up measurements are still part of the analysis. When applying longitudinal analysis of covariance (method 1), individuals with only a baseline measurement are not part of the analysis. Although some researchers claim that the repeated measures analysis takes into account the differences between the groups at baseline, there is actually no adjustment for the baseline differences. Therefore, an alternative to model 2 is developed in which the treatment variable is not part of the model, but its interaction with time still is (equations (2c) and (2d)).

$$Y_t = \beta_0 + \beta_1 time + \beta_2 time \times X \tag{2c}$$

$$Y_t = \beta_0 + \beta_1 dummytime_1 + \beta_2 dummytime_2 + \beta_3 dummytime_1 \times X + \beta_4 dummytime_2 \times X \tag{2d}$$

Because the treatment variable is not in the model, the baseline values for both groups are assumed to be equal and are reflected in the intercept of the model (i.e. β_0). The treatment effects can be directly obtained from the regression coefficients for the interactions between the treatment variable and time (the overall treatment effect over time; β_2 in equation (2c)) or between the treatment variable and the two dummy variables for time (treatment effect at the two time-points; β_3 and β_4 in equation (2d)).

2.1.3. Method 3: Analysis of changes

In the third method, not the actual values at the different time-points are modelled, but the changes between the baseline measurement and the first follow-up measurement and between the baseline measurement and the second follow-up measurement (equation (3a)).

$$Y_t - Y_{t0} = \beta_0 + \beta_1 X \tag{3a}$$

Although, it is sometimes suggested that the analysis of changes takes into account the difference at baseline, this is not the case and therefore this method can also be performed with an adjustment for the baseline value of the outcome variable (equation (3b)).

$$Y_t - Y_{t0} = \beta_0 + \beta_1 X + \beta_2 Y_{t0} \tag{3b}$$

As in method 1, the model can be extended with time and the interaction between the treatment variable and time to estimate the effect of the intervention at the different follow-up measurements (equations

Table 3
Data structure needed to perform the analyses described in method 3.

Id	Outcome	time	treatment	baseline
1	$Y_{t1} - Y_{t0}$	0	1	Y_{t0}
1	$Y_{t2} - Y_{t0}$	1	1	Y_{t0}

(3c) and (3d)).

$$Y_t - Y_{t0} = \beta_0 + \beta_1 X + \beta_2 time + \beta_3 X \times time \quad (3c)$$

$$Y_t - Y_{t0} = \beta_0 + \beta_1 X + \beta_2 Y_{t0} + \beta_3 time + \beta_4 X \times time \quad (3d)$$

The overall treatment effect and the treatment effects at the two follow-up measurements can be obtained in the same way as been described for the longitudinal analysis of covariance (method 1). Table 3 shows the structure of the data used to estimate the parameters of the analysis of changes.

2.2. Example dataset

The example dataset is taken from an intervention study in which the effectiveness of a long-term homocysteine-lowering treatment with folic acid plus pyridoxine in reducing systolic blood pressure was evaluated [12]. In this 2-year, randomised, placebo-controlled trial, a baseline measurement and two follow-up measurements (after 1 and after 2 years) were performed. At each time-point systolic blood pressure was measured four times and the average value was used in the analysis.

2.3. Statistical analysis

All statistical analyses were performed with linear mixed models in STATA (version 14). Linear mixed models was used because in all methods, an adjustment should be made for the dependency of the repeated observations within the individual. This adjustment was performed by adding a random intercept to the model [13]. In the illustration, the differences between the groups and the 95% confidence intervals were calculated.

3. Results

Table 4 shows descriptive information regarding the example used in the present study. There is a difference between the baseline values of the two groups, i.e. the treatment group has lower systolic blood pressure values compared to the control group. Furthermore, the systolic blood pressure of the treatment group decreases from baseline to the first follow-up measurement and there is a smaller decrease between the first to the second follow-up. For the control group there is a very small decrease from baseline to the first follow-up and a slightly bigger decrease from the first to the second follow-up measurement.

Table 5 shows the results of the different methods to estimate the overall treatment effect (i.e. the treatment effect on average over time) in the example RCT, while Table 6 shows the estimated treatment effects separately for the two follow-up measurements.

The results of the longitudinal analysis of covariance (method 1) show an overall treatment effect of -3.7 mmHg; so on average over time the treatment group has a 3.7 mmHg lower systolic blood pressure

Table 4
Descriptive information[a] regarding the example dataset.

	Baseline	T1	T2
Treatment	126.5 (12.5); n = 68	122.6 (11.5); n = 63	121.6 (12.3); n = 59
Control	130.7 (17.6); n = 71	130.1 (17.0); n = 67	127.2 (14.4); n = 60

[a] Mean systolic blood pressure and SD between brackets.

Table 5
Overall treatment effect estimated with different methods.

Equation	Method	Overall treatment effect
(1a)	Longitudinal analysis of covariance	-3.7 (-6.8 to -0.6)[a]
(2a)	Repeated measures[b]	-6.2 (-10.7 to -1.7)[a]
(2c)	Repeated measures without treatment[b]	-3.5 (-6.6 to -0.3)[a]
(3a)	Analysis of changes (not adjusted)	-1.9 (-5.5 to 1.8)
(3b)	Analysis of changes (adjusted)	-3.7 (-6.8 to -0.6)[a]

[a] Statistically significant at α = 0.05.
[b] Time recoded to 0 for the baseline measurement and 1 for both follow-up measurements.

Table 6
Treatment effects at the two follow-up measurements estimated with different methods.

Equation	Method	Treatment effect	
		first follow-up	second follow-up
(1b)	Longitudinal analysis of covariance	-4.6 (-8.2 to -0.95)[a]	-2.7 (-6.4 to 1.1)
(2b)	Repeated measures	-7.1 (-12.0 to -2.2)[a]	-5.2 (-10.2 to -0.19)[a]
(2d)	Repeated measures without treatment	-4.3 (-8.0 to 0.59)	-2.4 (-6.3 to 1.4)
(3b)	Analysis of changes (not adjusted)	-2.7 (-6.9 to 1.4)	-0.83 (-5.1 to 3.4)
(3d)	Analysis of changes (adjusted)	-4.6 (-8.2 to -0.95)[a]	-2.7 (-6.4 to 1.1)

[a] Statistically significant at α = 0.05.

compared to the control group. At the two follow-up measurements the longitudinal analysis of covariance revealed treatment effects of -4.6 and -2.7 mmHg respectively. The magnitude of the effect estimated with longitudinal analysis of covariance is more or less expected given the difference at baseline between the groups, the influence of regression to the mean and the observed mean values of systolic blood pressure at the two follow-up measurements.

From the results it can further be seen that with the repeated measures analysis (method 2), no adjustment is made for the baseline differences. Basically the overall effect and the effects at the two follow-up measurements are comparable to the observed differences in systolic blood pressure between the groups. This is, however, an overestimation of the treatment effect. This overestimation is due to: 1) not take into account regression to the mean, 2) the fact that the control group has higher blood pressure values at baseline and 3) the fact that blood pressure decreases over time.

When the treatment group variable is not in the model, actually an adjustment is made for the baseline differences. However, the results are slightly different from the results obtained from the longitudinal analysis of covariance (method 1).

In the present example, the analysis of changes (method 3) lead to an underestimation of the treatment effect. Because a decrease is harder to achieve in the treatment group (due to the lower values at baseline), the observed decrease in the treatment group should be weighted more heavily than the observed decrease in the control group. Analysing the change between baseline and the two follow-up measurements ignores this and therefore, this analysis lead to an underestimation of the treatment effect. When in the analysis of changes an adjustment is made for the baseline value, the results are exactly the same as the results of the longitudinal analysis of covariance (method 1).

4. Discussion

The present educational paper contains an explanation of different methods to estimate treatment effects in RCTs. Besides that, the

---

Longitudinal analysis of covariance (method 1):

$Y_t = \beta_0 + \beta_1 X + \beta_2 Y_{t0}$

Analysis of changes (method 3):

$Y_t - Y_{t0} = \beta_0 + \beta_1 X + \beta_2 Y_{t0}$

$Y_t = \beta_0 + \beta_1 X + \beta_2 Y_{t0} + Y_{t0}$

$Y_t = \beta_0 + \beta_1 X + (1 + \beta_2) Y_{t0}$

When the equation of analysis of changes is rewritten to define the outcome Y_t, only the regression coefficient of Y_{t0} changes by a value of 1, the coefficient of the treatment effect and the intercept remain the same. So whether Y_t or Y_t - Y_{t0} is being used as an outcome, all regression coefficients will remain the same except for the coefficient of Y_{t0} which will be higher by a value of 1 for the model with outcome Y_t compared to the model with outcome Y_t - Y_{t0}.

Fig. 1. Mathematical equivalence between longitudinal analysis of covariance and the analysis of changes with an adjustment for baseline differences.

different methods are used to estimate treatment effects of an example RCT aiming to decrease systolic blood pressure. Due to the observed differences in the outcome variable at baseline, the results of the different methods were totally different.

Because in the example dataset, the treatment group has a lower mean blood pressure at baseline, regression to the mean tend to increase blood pressure for the treatment group and tend to decrease blood pressure for the control group. Because of that, the treatment effect estimated with the analysis of changes (method 3), leads to an underestimation of the treatment effect, while the treatment effect estimated with the repeated measures analysis (method 2) leads to an overestimation of the treatment effect. Regarding the adjustment for the baseline value, it does not matter whether the outcome variable is the absolute value at the different follow-up measurements (i.e. longitudinal analysis of covariance; method 1) or the changes between the two follow-up measurements and the baseline measurement (i.e. analysis of changes; method 3); the effect estimates are exactly the same in both methods. In fact, there is a mathematical equivalence between the two methods leading to the same estimation of the treatment effect (see Fig. 1).

Although the general idea is the same, the results of repeated measures analysis without the treatment variable in the model (equations (2c) and (2d)) slightly differed from the results of the longitudinal analysis of covariance (method 1). The advantage of the repeated measures analysis (method 2) is that also subjects with only a baseline measurement are included in the analysis, which is not the case in the longitudinal analysis of covariance (method 1). So, in the present example the two analyses are based on a slightly different population. However, also when the method is used in a dataset without any missing data, the results of the two methods are not exactly the same. This is partly caused by the adjustment for the dependency of the repeated observations within the individual by adding a random intercept to the model. In the repeated measures analysis (method 2) using all three measurements as outcome, this random intercept variance is much higher than in the longitudinal analysis of covariance (method 1). In the latter, part of the random intercept variance is explained by the baseline value of the outcome which is added as an independent variable the model.

Longitudinal analysis of covariance (method 1): Analysis of changes (method 3):

$Y_t = \beta_0 + \beta_1 X + \beta_2 Y_{t0} Y_t - Y_{t0} = \beta_0 + \beta_1 X + \beta_2 Y_{t0}$

$Y_t = \beta_0 + \beta_1 X + \beta_2 Y_{t0} + Y_{t0}$

$Y_t = \beta_0 + \beta_1 X + (1 + \beta_2) Y_{t0}$

When the equation of analysis of changes is rewritten to define the outcome Y_t, only the regression coefficient of Y_{t0} changes by a value of 1, the coefficient of the treatment effect and the intercept remain the same. So whether Y_t or Y_t - Y_{t0} is being used as an outcome, all regression coefficients will remain the same except for the coefficient of Y_{t0} which will be higher by a value of 1 for the model with outcome Y_t compared to the model with outcome Y_t - Y_{t0}.

4.1. Statistical significance of baseline differences

It is often argued that an adjustment for baseline differences is only necessary when the difference between the groups at baseline is statistically significant. This is, however, a huge misunderstanding. Basically, the baseline value of the outcome variable can be seen as a confounder in the estimation of the treatment effect. A variable is considered to be a confounder when it is related to both the independent and the dependent variable in the model. Because in an RCT, the baseline value of the outcome is highly related to the follow-up measurements of the same variable, even a small difference in baseline value of the outcome between the two groups can have a (strong) confounding effect. Therefore, it is advised always to adjust for the baseline value of the outcome variable irrespective whether the difference is significant or not. The issue of statistical significance also holds for the adjustment for other covariates. Although the adjustment for other covariates is less important than the adjustment for the baseline value of the outcome, it can still be important to consider adjustment for other covariates [14]. When a covariate is related to the outcome and when that covariate differs between the two groups, the particular covariate is considered to be a confounder in the estimation of the treatment effect. Again, note that it is not necessary that the covariate is significantly related to the outcome or that the covariate is significantly different between the two groups. Significance does not play an important role in the magnitude of the influence of the particular covariate. It is therefore of no use to statistically test for baseline differences between the treatment and the control group. This testing nonsense has been noticed by other authors as well [15,16].

4.2. Dichotomous outcomes

When the outcome variable in an RCT is dichotomous, (longitudinal) logistic regression analysis is used to estimate treatment effects. With dichotomous outcomes, mostly an adjustment for baseline differences in the outcome is not necessary, because at baseline mostly all individuals are either scoring 1 or 0 (depending on the coding of the particular outcome). Suppose that one wants to estimate the effect of a new treatment against hypertension, in the source population all

subjects must have hypertension. In other words, they all have the same value of the outcome variable at baseline. When this is not the case, i.e. when there is a difference in the baseline dichotomous outcome between the treatment and the control group, the situation is slightly more complicated than described for continuous outcomes. This has to do with the fact that in (longitudinal) logistic regression analysis the effect estimate changes when a variable which is highly related to the outcome is added to the model. This change is irrespective of the difference in this variable between the two groups. So when the baseline values of the two groups are exactly the same and the baseline value is (highly) related to the outcome, the result of the unadjusted (longitudinal) logistic regression analysis will differ from the result of the adjusted (longitudinal) logistic regression analysis. This phenomenon is known as non-collapsibility [17–19] and arises from differences in the total variances between a logistic model with the adjustment of the particular variable and a logistic model without the adjustment. Basically the total variance is the summation of explained and unexplained variance. When a covariate is added to a linear regression model, the unexplained variance decreases while the explained variance increases with the same amount. However, in a logistic model, the unexplained variance is a fixed number. So when a covariate that is related to the outcome is added to a logistic model which only contains the treatment variable, the total variance will increase. Because of this increased variance it is often said that, adding a variable to the logistic model that is related to the outcome changes the scale on which the regression coefficients must be interpreted and therefore, they cannot be compared to each other.

This phenomenon, which is not known by most researchers, is illustrated in Fig. 2. In this illustration it can be seen that there is no difference in baseline values between the treatment and control group. The crude treatment effect (i.e. without adjusting for the baseline value) gives an odds ratio (OR) of 5. It can also be seen that the baseline value is highly related to the outcome variable (the corresponding OR also equals 5). Because the baseline value is exactly the same for both the intervention and control group an adjustment for the baseline value should not change the estimation of the treatment effect. However, when a logistic regression analysis is performed adjusting for the baseline value the adjusted OR will be around 6.7; much higher than the expected OR of 5.

4.3. Observational studies

It should be realised that the adjustment for baseline differences between groups is only necessary in RCTs. In observational studies differences at the first measurement between groups mostly reflect real differences. In observational studies it is, therefore, not advised to adjust for the differences between the groups, because then the real differences are neglected [20,21]. Surprisingly, in observational studies the first measurement is also often referred to as the baseline measurement. To avoid confusion and misunderstanding regarding the adjustment for baseline values within observational studies, it is probably better to talk about the 'first measurement' of an observational study instead of the baseline measurement. The term baseline measurement should be restricted to RCTs.

4.4. Assumptions

It should be realised that the results of the different methods depends partly on the assumptions made. For instance, the comparison between the different methods only holds for situations when the variances at the follow-up measurements are more or less the same for both groups and when the relationship between the baseline value and the follow-up measurements is more or less equal for both groups. When this is not the case, it is, for instance, shown that the standard errors obtained from a longitudinal analysis of covariance (method 1) are slightly underestimated [5]. Besides that, the comparison between the different methods is related to the absolute differences in the outcome variable between the groups. Although this is by far the mostly used effect estimate, it is not known how the different methods behave when the effect of the outcome is more relative, i.e. when the effect of the intervention is proportional to the individual baseline value. Because it is slightly beyond the scope of the present paper, for more technical details about the assumptions for the different methods one is referred to other papers [4–8,22].

It was already mentioned that the adjustment for baseline differences is only appropriate when the expected values of the baseline value is the same for both groups. This is the case in an RCT, but not in a non-randomised trial or an observational study.

Finally, in the mixed model analysis, we only added a random

a) no differences between intervention and control (OR (intervention/control) = 1)

	Intervention	Control
Baseline = 0	30	30
Baseline = 1	30	30

b) crude treatment effect (OR (intervention/control) = 5)

	Intervention	Control
Follow-up = 0	10	30
Follow-up = 1	50	30

c) strong relationship between baseline and follow-up (OR (baseline = 1/baseline = 0) = 5)

	Follow-up = 1	Follow-up = 0
Baseline = 1	30	30
Baseline = 0	10	50

Fig. 2. Illustration of the problem of non-collapsibility of the OR. a) no differences between intervention and control (OR (intervention/control) = 1).

intercept to take into account the dependency of the repeated observations within the subject. We decided not to add random slopes for the time variable to the models, because firstly time was not present in all models and secondly, when random slopes are added for the time dummy variables, it often leads to non-converging models.

5. Conclusion

To estimate a treatment effect in an RCT, the analysis has to be adjusted for the baseline value of the outcome variable. A proper adjustment is not achieved by performing a regular repeated measures analysis (method 2) or by the regular analysis of changes (method 3). It is advised to use either a longitudinal analysis of covariance (method 2) or a repeated measures analysis without the treatment variable, but with the interaction between the treatment variable and time in the model.

Funding

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.conctc.2018.03.008.

References

[1] G.M. Fitzmaurice, N.M. Laird, J.H. Ware, Applied Longitudinal Data Analysis, Wiley, Hoboken, NJ, 2004.

[2] J. Jamieson, Analysis of covariance (ANCOVA) with different scores, Int. J. Psychother. 52 (2004) 277–283.

[3] N. Mayer-Hamblett, R.A. Kronmal, Improving the estimation of change form baseline in a continuous outcome measure in a clinical trial setting, Contemp. Clin. Trials 26 (2005) 2–16.

[4] G.J.P. Breukelen, ANCOVA verus change from baseline had more power in randomized studies and more bias in nonrandomized studies, J. Clin. Epidemiol. 59 (2006) 920–925.

[5] B. Winkens, G.J.P. Breukelen, H.J.A. Schouten, M.P.F. Berger, Randomized clinical trials with a pre- and a post-treatment measurement: repeated measures versus ANCOVA models, Contemp. Clin. Trials 28 (2007) 713–719.

[6] Y. Petscher, C. Schatschneider, A simulation study on the performance of the simple difference and covariance-adjusted scores in randomized experimental designs, J. Educ. Meas. 48 (2011) 31–43.

[7] S. Senn, Regression to the mean, Stat. Meth. Med. Res. 6 (1997) 99–102.

[8] S.F. Chan, P. Macaskill, L. Irwig, S.D. Walter, Adjustment for baseline measurement error in randomized controlled trial induces bias, Contr. Clin. Trials 25 (2004) 408–416.

[9] J.L. Cummings, R.E. Tractenberg, A. Gamst, L. Teri, D. Masterman, L.J. Thal, Regression to the mean: implications for clinical trials of psychotropic agents in dementia, Curr. Alzheimer Res. 1 (2004) 323–328.

[10] A.G. Barnett, Pols JC. van der, A.J. Dobson, Regression to the mean: what it is and how to deal with it, Int. J. Epidemiol. 34 (2005) 215–230.

[11] P.C. Austin, A. Manca, M. Zwarenstein, D.N. Juurlink, M.B. Stanbrook, A substantia land confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals, J. Clin. Epidemiol. 63 (2010) 142–153.

[12] R.A. van Dijk, J.A. Rauwerda, M. Steyn, J.W. Twisk, C.D. Stehouwer, Long-term homocysteine-lowering treatment with folic acid plus pyridoxine is associated with decreased blood pressure but not with improved brachial artery endothelium-dependent vasodilation or carotid artery stiffness: a 2-year, randomized, placebo-controlled trial, Arterioscler. Thromb. Vasc. Biol. 21 (12) (2001) 2072–2079.

[13] J.W.R. Twisk, Applied Longitudinal Data Analysis for Epidemiology, second ed., Cambridge University Press, UK, 2013.

[14] B.C. Kahan, The risk and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies, Trials 15 (2014) 139.

[15] M.R. de Boer, W.E. Waterlander, L.D. Kuijper, I.H. Steenhuis, J.W. Twisk, Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate, Int. J. Behav. Nutr. Phys. Activ. 12 (2015 Jan 24) 4.

[16] R.L. Petterson, M. Tran, J. Koffel, S.D. Stovitz, Statistical testing of baseline differences in sports medicine RCT's: a systematic evaluation, BMJ Open Sports Exerc. Med. (2017) 3.

[17] S.C. Newman, Commonalities in the classical, collapsibility and counterfactual concepts in confounding, J. Clin. Epidemiol. 57 (2004) 325–329.

[18] S. Greenland, J.M. Robins, Identifiability, exchangeability and confounding revisited, Epidemiol. Perspect. Innovat. 6 (2009) 4.

[19] M.A. Hernan, D. Clayton, N. Keiding, The Simpson's paradox unravelled, Int. J. Epidemiol. 40 (2011) 780–785.

[20] H.C. Boshuizen, RE: twisk and Proper: evaluation of the results of a randomized controlled trial: how to define changes between baseline and follow-up, J. Clin. Epidemiol. 58 (2005) 209–210.

[21] J. Twisk, K. Proper, Is analysis of covariance the most appropriate way to analyse changes in randomized controlled trials, J. Clin. Epidemiol. 58 (2005) 211–212.

[22] M.J. Egger, M.L. Coleman, J.R. Ward, J.C. Reading, H.J. Williams, Uses and abuses of analysis of covariance in clinical trials, Contr. Clin. Trials 6 (1985) 12–24.