

The analysis of randomised controlled trial data with more than one follow-up measurement. A comparison between different approaches

Jos W. R. Twisk · Wieke de Vente

Received: 20 February 2008 / Accepted: 18 July 2008 / Published online: 19 August 2008
© The Author(s) 2008. This article is published with open access at Springerlink.com

Abstract When more than one follow-up measurement is analysed in a randomized controlled trial, there is no consensus how to analyse the overall intervention effect in a proper way. Mostly, longitudinal analysis of covariance is used, because with this method a correction is made for possible regression to the mean. However, in this paper it is shown that this method (mostly) leads to an overestimation of the intervention effect. A possible solution is the use of autoregression, although this does not seem to be the best solution, because it leads to an overcorrection. Due to these flaws, in this paper a new approach is introduced in which a correction for the baseline value is made for the first follow-up, but no correction is made for the remaining follow-up measurements.

Keywords Longitudinal studies · Analysis of covariance · Autoregression · Regression to the mean · Methodology

Abbreviations

GEE Generalised estimating equations
RCT Randomised controlled trial
SBP Systolic blood pressure

Background

When a continuous outcome variable from a randomised control trial (RCT) is analysed, it is recommended to use analysis of covariance in order to obtain a valid estimation of the intervention effect [1–4]. The idea behind using analysis of covariance is that a correction is made for regression to the mean. Regression to the mean at follow-up is expected to occur when the mean baseline values of the intervention and control group differ from one another. Essentially, differences in mean baseline values between the intervention and control group are not expected, as these groups come from the same population. If differences at baseline do indeed occur, these differences are due to chance (i.e. random fluctuations and/or measurement error). Correction for regression to the mean using analysis of covariance can be achieved by addition of the baseline value as a covariate in an analysis in which the follow-up measurement is the outcome variable and a group allocation is the independent variable. Not correcting for baseline differences can lead to either over- or underestimation of the estimated intervention effect [4].

In most RCT's, however, more than one follow-up measurement is conducted. With the availability of sophisticated statistical analysis, such as generalised estimating equations (GEE-analysis) and mixed effects modelling, it is common to analyse the whole development of the outcome variable over time in one analysis, instead of analysing separate measurements. These longitudinal designs are

J. W. R. Twisk
Department of Methodology and Applied Biostatistics,
Institute of Health Sciences, Vrije Universiteit,
Amsterdam, The Netherlands

J. W. R. Twisk (✉)
Department of Clinical Epidemiology and Biostatistics,
Vrije Universiteit medical centre, de Boelelaan 1118,
1081 HV Amsterdam, The Netherlands
e-mail: jwr.twisk@vumc.nl

W. de Vente
Department of Educational Sciences,
University of Amsterdam, Amsterdam,
The Netherlands

frequently analysed using a longitudinal analysis of covariance [5–8]. However, it is questionable whether this is the right approach, because the result of this analysis is highly influenced by the first period of the study [9]. Therefore, it is sometimes suggested to use a so-called autoregressive approach [9–11]. In this approach, the outcome variable at a certain time-point is corrected for the value of the outcome variable one time-point earlier. So, the measurement at the first follow-up measurement is corrected for the baseline value of the outcome variable, while the outcome variable at the second follow-up measurement is corrected for the value of the outcome variable at the first follow-up measurement, etc. Although this autoregression approach seems to be an adequate solution, a new problem arises. The correction for differences in the outcome variable at baseline is performed because it is assumed that the groups to be compared are equal at baseline. At the first follow-up measurement the situation is different, though, because differences between the groups can now be caused by the fact that one group received the intervention and the other group did not [12, 13]. So except for the first follow-up measurement, correction for previous measurements does not make sense and is therefore not recommended [3]. Instead, we are searching for an analytic approach that combines a correction for the baseline value for the outcome variable at the first follow-up measurement, with a procedure that does not correct for any values of the outcome variable at the next follow-up measurements.

The purpose of this paper is, therefore, to propose such an approach and to compare the results of this new combination approach with the approaches that are normally used in longitudinal analyses of data from RCT's.

Methods

Two examples will be used to illustrate the different approaches. The data used in the first example are simulated from an RCT evaluating a new intervention to reduce blood pressure. 78 subjects were randomised into the control group and 74 subjects received the intervention. Besides the baseline measurement, two follow-up measurements were conducted. Figure 1 shows the development over time in the outcome variable systolic blood pressure for both groups.

The data used in the second example are simulated from an RCT investigating an intervention aiming to increase physical activity. 118 subjects were randomised into the control group and 99 subjects received the intervention. In the second example, the subjects were measured at baseline and at three follow-up measurements. Figure 2 shows the development of the outcome variable physical activity over the four measurements in both groups.

Systolic blood pressure

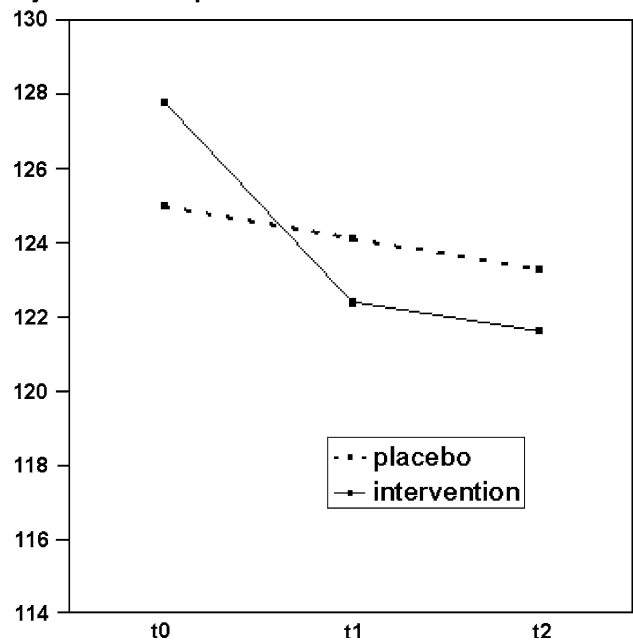


Fig. 1 Development of systolic blood pressure over time for both the intervention and the control group (example 1)

Physical activity

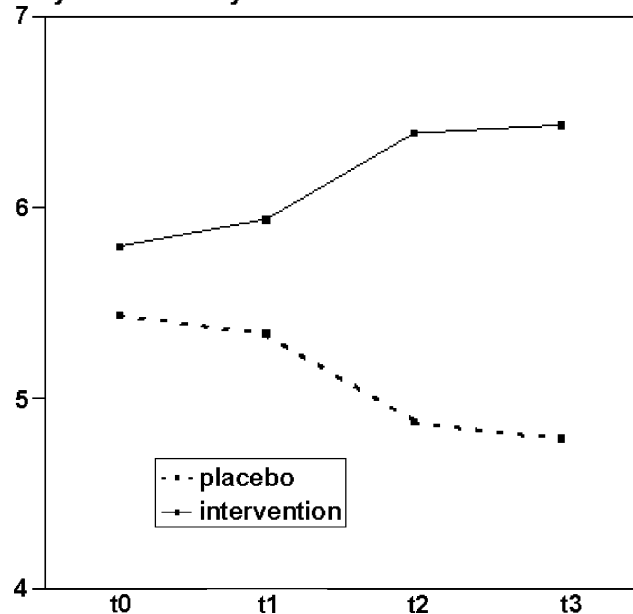


Fig. 2 Development of physical activity over time for both the intervention and the control group (example 2)

Analyses

In this paper, five different approaches are compared. In the first approach, possible regression to the mean is ignored and the changes between subsequent measurements are

used as outcome variables. This approach is referred to as the longitudinal analysis of changes.

Second, analysis of covariance is used. By using longitudinal analysis of covariance, the outcome variable at all follow-up measurements is corrected for the baseline value (formula 1).

$$y_{it} = b_0 + b_1 \text{int}_i + b_2 y_{i0} + \varepsilon_{it} \tag{1}$$

Where: y_{it} = outcome variable for subject i at time t ; b_0 = intercept; int = intervention variable for subject i ; b_1 = regression coefficient for the intervention variable; y_{i0} = baseline value of the outcome variable subject i ; b_2 = regression coefficient for the baseline value and ε_{it} = ‘error’ for subject i at time t .

The third approach is autoregression. In this approach the outcome variable at the different follow-up measurements is corrected for the value of the outcome variable one time-point earlier (formula 2).

$$y_{it} = b_0 + b_1 \text{int}_i + b_2 y_{it-1} + \varepsilon_{it} \tag{2}$$

Where: y_{it} = outcome variable for subject i at time t ; b_0 = intercept; int = intervention variable for subject i ; b_1 = regression coefficient for the intervention variable; y_{it-1} = value of the outcome variable for subject i at $t - 1$; b_2 = regression coefficient for the outcome variable at $t - 1$ and ε_{it} = ‘error’ for subject i at time t .

The fourth and the fifth approach are combined approaches in line with our aim, namely to correct for regression to the mean by *only* adjusting the first follow-up measurement of the outcome variable for its baseline value. The fourth approach is a combination of the ‘residual change’ method, first described by Blomquist [14], and ‘normal’ change scores. The first step in the residual change method is to perform a linear regression analysis between the follow-up measurement and the baseline value. The second step is to calculate the difference between the observed value at the follow-up measurement and the predicted value (i.e. the predicted value by the earlier mentioned regression analysis). This difference is called the ‘residual change’. The third step is to use these residual change scores as the outcome variable in a linear regression analysis in which the effect of the intervention is analysed. The advantage of this method is that no further correction has to be performed, because the correction for the baseline value is already reflected in the outcome variable. In the combination approach to analyse longitudinal data, the ‘residual change’ score is used as outcome for the first follow-up measurement, while the ‘normal’ change scores are used as outcome for the other follow-up measurements. This approach will be further referred to as the ‘residual change’ combination.

The fifth approach is a combination of analysis of covariance and ‘normal’ change scores. In this approach,

the change between baseline and the first follow-up is corrected for the baseline value. For the remaining follow-up measurements, the change scores between subsequent measurements can be used without any correction. However, due to the fact that the first change has to be corrected for the baseline value, a correction factor for the remaining change scores must be added to the longitudinal regression model. Because this correction is only necessary for computational reasons without having any influence on the estimated regression coefficient, a correction must be made for the mean value of the outcome variable at the beginning of the specific period over which the change score is calculated. This approach will be further referred to as the ‘analysis of covariance’ combination.

For all approaches, the intervention effect is estimated with GEE-analysis and all GEE-analyses were performed with STATA (version 10).

Results

Table 1 shows the results of the different approaches to analyse the effect of the intervention in the blood pressure example. Analysis of covariance gives the largest intervention effect. In the intervention group, systolic blood pressure is (on average) 3.1 mmHg lower than in the control group. This intervention effect is also statistically significant ($P = 0.04$). All other approaches showed non-significant intervention effects. The effect sizes are more or less similar; the ‘residual change’ combination approach resulted in the smallest intervention effect.

Table 2 shows the results of the different approaches to analyse the effect of the intervention in the physical activity example. Although all analyses show a significant intervention effect, there is a remarkable difference in magnitudes of these effects. As in the first example,

Table 1 Results (regression coefficients, 95% confidence intervals and P -values) of various analyses to estimate the effect of the intervention on blood pressure reduction (example 1)

Analysis	Regression coefficient	95% CI	P -value
Analysis of changes	-1.889	-3.964 to 0.185	0.074
Analysis of covariance	-3.089	-6.037 to -0.142	0.040
Autoregression	-1.836	-3.875 to 0.203	0.078
‘Residual change’ combination	-1.539	-3.318 to 0.240	0.090
‘Analysis of covariance’	-1.587	-3.339 to 0.165	0.076

Table 2 Results (regression coefficients, 95% confidence intervals and *P*-values) of various analyses to estimate the effect of the intervention on physical activity improvement (example 2)

Analysis	Regression coefficient	95% CI	<i>P</i> -value
Analysis of changes	0.427	0.363 to 0.492	<0.001
Analysis of covariance	0.993	0.813 to 1.173	<0.001
Autoregression	0.556	0.463 to 0.650	<0.001
'Residual change' combination	0.458	0.400 to 0.516	<0.001
'Analysis of covariance' combination	0.461	0.400 to 0.522	<0.001

longitudinal analysis of covariance shows the strongest intervention effect, while in this example the longitudinal analysis of changes resulted in the weakest intervention effect. The results of the two combination approaches are almost equal.

Discussion

In this paper, five different approaches to analyse data from a longitudinal RCT were compared to each other. The most remarkable finding regarding the three commonly used methods to analyse longitudinal data of randomised controlled trials was the fact that longitudinal analysis of covariance highly overestimated the intervention effects. This overestimation is due to the fact that in longitudinal analysis of covariance, each follow-up measurement is corrected for baseline. Hence, the (corrected) change between baseline and the first follow-up, the (corrected) change between baseline and the second follow-up, etc is added up. When, for instance, the (corrected) change between baseline and the second follow-up is considered, this also includes the (corrected) change over the first measurement period. This phenomenon is illustrated in Fig. 3. With longitudinal analysis of covariance for the placebo group, $[a1 + (a1 + a2) + (a1 + a2 + a3)]$ is compared with $[b1 + (b1 + b2) + (b1 + b2 + b3)]$ for the intervention group. So, the (corrected) change over the first period is counted three times. Overestimation of the intervention effect particularly occurs when the intervention has its main effect in the first period of the study. Actually, in intervention studies, this is typically the case. In addition, overestimation increases as the number of follow-up measurements increases.

The problem with longitudinal analysis of changes is that it ignores possible differences at baseline between the two groups. The larger the difference between the two groups at baseline, the more the validity of this analysis will be violated. Using the longitudinal analysis of changes

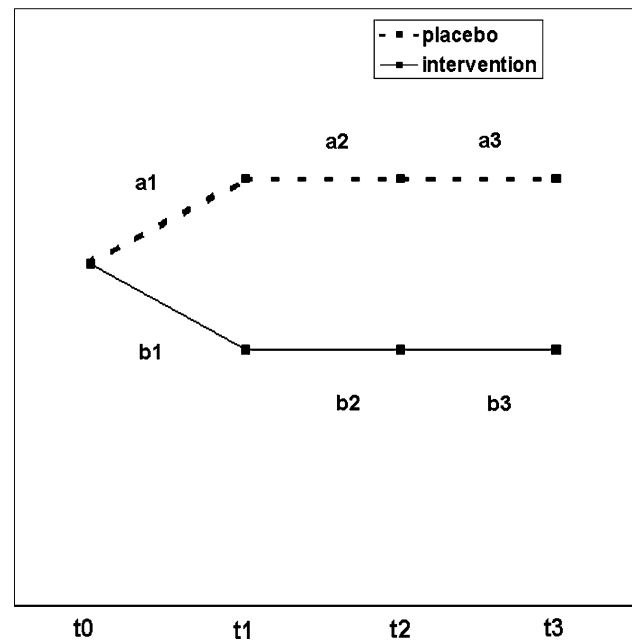


Fig. 3 Schematic illustration of the results of an RCT to show the 'problems' of longitudinal analysis of covariance

can either lead to an overestimation or an underestimation of the intervention effect. This depends on whether the intervention group is higher or lower than the control group at baseline and whether the intervention leads to a reduction or to an increase in the outcome variable. In the first example we have seen that the intervention group starts at a higher level and that the intervention led to a reduction in the outcome variable. In this situation the longitudinal analysis of changes resulted in an overestimation of the intervention effect. In the second example, we have seen that the intervention group also starts at a higher level, but the intervention results in an increase in the outcome variable. In this situation the longitudinal analysis of changes results in an underestimation of the intervention effect.

Regarding autoregression, as mentioned before, correcting follow-up measurements after the first one can be considered as overcorrection. Correcting the first follow-up for the baseline value is a correction for group differences due to chance and is thus necessary to obtain a valid estimate of the intervention effect. After the intervention started, however, the subjects do not come from the same population anymore, so differences at a follow-up measurement are not due to chance. Instead, they are probably due to the intervention. Therefore, a correction for follow-up measurements is not necessary for an adequate estimation of the intervention effect, and may even cause bias. Although autoregression can be considered as an overcorrection, it mostly leads to an overestimation of the intervention effect. This was also the case in the two examples described in this paper. Autoregression leads to

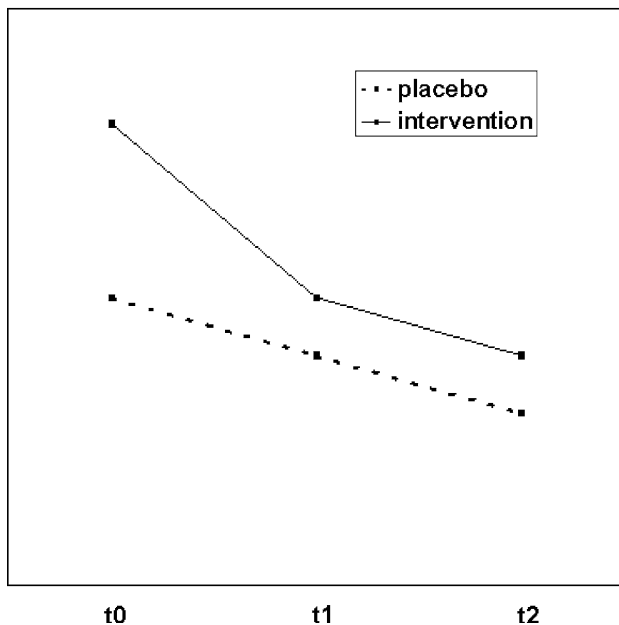


Fig. 4 Result of an RCT in which autoregression leads to an underestimation of the ‘real’ intervention effect

an underestimation of the intervention effect when, for instance, the intervention leads to a reduction in the outcome variable and the intervention group has a higher value than the control group over the whole follow-up period (see Fig. 4).

Combination approaches

In this paper, two combination approaches were proposed in which the first follow-up measurement was adjusted for differences at baseline and in which the next follow-up measurement was not corrected for either baseline or the earlier follow-up measurement. Although the ‘residual change’ score is commonly assumed to be equal to analysis of covariance, it is known that the use of ‘residual change’ scores always leads to lower intervention effects, compared to analysis of covariance [15, 16]. This can also be seen in Tables 1 and 2, where the ‘residual change’ combination leads to lower intervention effects compared to the ‘analysis of covariance’ combination approach. Although this underestimation of the use of the ‘residual change’ score is generally small, Forbes and Carlin [16] showed that it can be non-trivial when the number of participants reduces, and/or baseline differences increase, and/or the ‘true’ intervention effect gets larger.

Because of this drawback of the ‘residual change’ scores, the ‘analysis of covariance’ combination approach is introduced in the present study as a very nice alternative. In the presented analyses, the change between baseline and the first follow-up measurement was used as first outcome in the

longitudinal analyses. It is also possible to use the value of the outcome variable at the first follow-up measurement as first outcome variable in the longitudinal analyses. When both are corrected for baseline, the estimated effect of the intervention will be exactly the same [9].

Dichotomous outcome variables

In the example datasets, continuous outcome variables were considered. In RCT’s where the outcome variable is dichotomous, there is generally no problem of regression to the mean. This has to do with the fact that normally no baseline differences regarding the outcome variable exist between the intervention and control group. Regarding the two examples used in the present study, for instance, hypertension could have been the outcome variable instead of systolic blood pressure. At the beginning of the study all patients must have hypertension, because otherwise they could not be included in the study. Likewise, participant’s level of physical activity in comparison to the recommended level could be the outcome, rather than the exact level of physical activity. One of the inclusion criteria would have been to have an activity level below the recommended level and again, no group differences would be present, hence, no baseline correction would be necessary.

Conclusion

This paper shows that the ‘analysis of covariance’ combination approach, in which only the first follow-up measurement is adjusted for baseline, is the best way of estimating the intervention effect in longitudinal RCT’s. This approach is more adequate than the autoregression approach or the longitudinal analysis of changes to analyse longitudinal data of an RCT. More importantly, though, in situations when two or more follow-up measurements are analysed together, we strongly advice against the commonly used longitudinal analysis of covariance, since it leads to considerable overestimation of the intervention effect. When only one follow-up measurement is analysed, analysis of covariance is still appropriate.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Senn S. Repeated measures in clinical trials: analysing using mean summary statistics and its implications for design. *Stat Med.* 1994;13:197–8. doi:10.1002/sim.4780130210.

2. Vickers A, Altman D. Analysing controlled trials with baseline and follow-up measurements. *BMJ*. 2001;323:1123–4. doi:[10.1136/bmj.323.7321.1123](https://doi.org/10.1136/bmj.323.7321.1123).
3. Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal analysis*. Hoboken, New Jersey, USA: Wiley; 2004.
4. Twisk J, Proper K. Evaluation of the results of a randomised controlled trial: how to define changes between baseline and follow-up. *J Clin Epidemiol*. 2004;57:223–8. doi:[10.1016/j.jclinepi.2003.07.009](https://doi.org/10.1016/j.jclinepi.2003.07.009).
5. Simon GE, Ludman EJ, Bauer MS, Unützer J, Operskalski B. Long-term effectiveness and cost of a systematic care program for bipolar disorder. *Arch Gen Psychiatry*. 2006;63:500–8. doi:[10.1001/archpsyc.63.5.500](https://doi.org/10.1001/archpsyc.63.5.500).
6. Giesen-Bloo J, van Dyck R, Spinhoven P, van Tilburg W, Dirksen C, van Asselt Kremers T, et al. Outpatient psychotherapy for borderline personality disorder. *Arch Gen Psychiatry*. 2006;63:649–58. doi:[10.1001/archpsyc.63.6.649](https://doi.org/10.1001/archpsyc.63.6.649).
7. Rowe M, Bellamy C, Baranoski M, Wieland M, O'Connell MJ, Benedict P, et al. A peer-support, group intervention to reduce substance use and criminality among persons with severe mental illness. *Psychiatr Services*. 2007;58:955–61. doi:[10.1176/appi.ps.58.7.955](https://doi.org/10.1176/appi.ps.58.7.955).
8. Goodyer I, Dubicka B, Wilkinson P, Kelvin R, Roberts C, Byford S, et al. Selective serotonin reuptake inhibitors (SSRIs) and routine specialist care with and without cognitive behaviour therapy in adolescents with major depression: randomised controlled trial. *BMJ*. 2007;335:142–9. doi:[10.1136/bmj.39224.494340.55](https://doi.org/10.1136/bmj.39224.494340.55).
9. Twisk JWR. *Applied longitudinal data analysis for epidemiology. A practical guide*. Cambridge, UK: Cambridge University Press; 2003.
10. Rosner B, Munoz A, Tager I, Speizer F, Weiss S. The use of an autoregressive model for the analysis of longitudinal data in epidemiologic studies. *Stat Med*. 1985;4:457–67. doi:[10.1002/sim.4780040407](https://doi.org/10.1002/sim.4780040407).
11. Twisk JWR. Different Statistical models to analyse epidemiological observational longitudinal data. *Int J Sports Med*. 1997;18(Suppl3):S216–24. doi:[10.1055/s-2007-972718](https://doi.org/10.1055/s-2007-972718).
12. Boshuizen HC. Re: Twisk and Proper: evaluation of the results of a randomised controlled trial: how to define changes between baseline and follow-up. *J Clin Epidemiol*. 2005;58:209–10. doi:[10.1016/j.jclinepi.2004.06.009](https://doi.org/10.1016/j.jclinepi.2004.06.009). Letter.
13. Twisk J, Proper K. *J Clin Epidemiol*. 2005;58:211–2. doi:[10.1016/j.jclinepi.2004.06.010](https://doi.org/10.1016/j.jclinepi.2004.06.010). Author reply.
14. Blomquist N. On the relation between change and initial value. *JASA*. 1977;72:746–7499.
15. Maxwell S, Delaney H, Manheimer J. ANOVA of residuals and ANCOVA: Correcting an illusion by using model comparisons and graphs. *J Educ Stat*. 1985;10:197–209. doi:[10.2307/1164792](https://doi.org/10.2307/1164792).
16. Forbes AB, Carlin JB. Residual change analysis is not equivalent to analysis of covariance. *J Clin Epidemiol*. 2005;58:540–1. doi:[10.1016/j.jclinepi.2004.12.002](https://doi.org/10.1016/j.jclinepi.2004.12.002).