

A comparison of procedures to correct for base-line differences in the analysis of continuous longitudinal data: a case-study

Peer-reviewed author version

VERBEKE, Geert; FIEUWS, Steffen; LESAFFRE, Emmanuel; Kato, B.S.; Foreman, M.D.; Broos, P.L.O. & Milisen, Koen (2006) A comparison of procedures to correct for base-line differences in the analysis of continuous longitudinal data: a case-study. In: JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES C-APPLIED STATISTICS, 55(1). p. 93-101.

DOI: 10.1111/j.1467-9876.2005.00531.x

Handle: <http://hdl.handle.net/1942/1571>

A comparison of procedures to correct for base-line differences in the analysis of continuous longitudinal data: a case-study

G. Verbeke, S. Fieuws and E. Lesaffre,
Katholieke Universiteit Leuven, Belgium

B. S. Kato,
St Thomas's Hospital, London, UK

M. D. Foreman
University of Illinois at Chicago, USA

and P. L. O. Broos and K. Milisen
Katholieke Universiteit Leuven, Belgium

[Received November 2004. Final revision June 2005]

Summary. The main advantage of longitudinal studies is that they can distinguish changes over time within individuals (longitudinal effects) from differences between subjects at the start of the study (base-line characteristics; cross-sectional effects). Often, especially in observational studies, subjects are very heterogeneous at base-line, and one may want to correct for this, when doing inferences for the longitudinal trends. Three procedures for base-line correction are compared in the context of linear mixed models for continuous longitudinal data. All procedures are illustrated extensively by using data from an experiment which aimed at studying the relationship between the post-operative evolution of the functional status of elderly hip fracture patients and their preoperative neurocognitive status.

Keywords: Base-line correction; Linear mixed model; Longitudinal data; Longitudinal trends

1. Introduction

In health sciences, studies are often designed to investigate changes in a specific parameter which is measured repeatedly over time in the participating subjects. Such studies are in contrast with cross-sectional studies where the response of interest is measured only once for each individual. As pointed out by Diggle *et al.* (1994), section 1.4, the main advantage of longitudinal studies is that they can distinguish changes over time within individuals (longitudinal effects) from differences between people in their base-line values (cross-sectional effects).

Often, especially in observational studies, much base-line variability between subjects exists and we wish to study longitudinal trends, correcting for these base-line differences. For example, Diggle *et al.* (1994), section 9.3, used longitudinal data on 250 children to investigate the evolution of the risk for respiratory infection, and its relationship to vitamin A deficiency. They

Address for correspondence: G. Verbeke, Biostatistical Centre, Katholieke Universiteit Leuven, Universitaire Ziekenhuizen St-Rafaël, Kapucijnenvoer 35, B-3000 Leuven, Belgium.
E-mail: geert.verbeke@med.kuleuven.be

adjusted for gender, season and age of subjects on entry to the study. Here the correction was done by including known base-line characteristics as covariates in the model. An alternative technique is the use of conditional linear mixed models (Verbeke *et al.*, 2001) which considers the base-line values as nuisance parameters which can be conditioned out before the estimation of the longitudinal trends. A third technique, which has also been proposed for the analysis of pretest–post-test data (Bonate, 2000) uses the first observed outcome as a covariate in a model for the subsequent observations. We shall call this technique ‘analysis of covariance’, thereby acknowledging the analogy with the more traditional linear regression models that are used for comparison of groups, correcting for covariates.

In Section 2, the motivating data set will be introduced and analysed by using a linear mixed model, not correcting for base-line differences. In Section 3, we present three procedures for base-line correction, and, in Section 4, we apply these methods to our data set. In Section 5, a general discussion will be given explaining the results that are observed in the data analysis. Finally, concluding remarks are given in Section 6.

2. Case-study: activities of daily living data

The motivation for this research was the analysis of data that have been collected as part of a doctoral research project in the Centre for Health Services and Nursing Research of the Katholieke Universiteit Leuven in Belgium. Details about the project can be found in Milisen (1999) and Milisen *et al.* (1998). The aim of this part of the project was to study the relationship between the post-operative evolution of the functional status of elderly hip fracture patients and their preoperative neurocognitive status. Functional status was measured on the first, fifth and 12th day post-operatively, using the Belgian version of the Katz index of activities of daily living (ADL) (Katz and Akpom, 1976; Delesie *et al.*, 1987), which is an ordinal score ranging from 6 to 24, with high ADL values indicating greater dependence on others by the subject. In total, we have data on 54 patients, 17 of whom were classified as having neurocognitive symptoms before the surgery. The individual profiles of all 54 subjects are shown in Fig. 1, for the neuropsychiatric and non-neuropsychiatric patients separately. Let $Y_i(t)$ now denote the ADL response taken for subject $i = 1, \dots, 54$, at time points $t = 1, 5, 12$. We then assume that $Y_i(t)$ can be modelled as

$$Y_i(t) = \begin{cases} \beta_{10} + b_{0i} + (\beta_{11} + b_{1i}) \ln(t) + \varepsilon_i(t) & \text{if neuropsychiatric,} \\ \beta_{20} + b_{0i} + (\beta_{21} + b_{1i}) \ln(t) + \varepsilon_i(t) & \text{if non-neuropsychiatric.} \end{cases} \quad (1)$$

The model assumes that, on a logarithmic timescale, the ADL score evolves linearly over time, with average intercepts β_{10} and β_{20} and average slopes β_{11} and β_{21} , for the neurocognitive patients and the patients without preoperative neurocognitive symptoms respectively. Further, every subject i is allowed to have its own intercept and slope, through the inclusion of the random intercepts b_{0i} and random slopes b_{1i} which are assumed to follow jointly a two-dimensional normal distribution with mean $\mathbf{0}$ and covariance matrix D . Finally, all error components $\varepsilon_i(t)$ are assumed to be independent of each other as well as of the random effects b_{0i} and b_{1i} , and normally distributed with mean 0 and variance σ^2 .

Note that the components $\beta_{10} + b_{0i}$ and $\beta_{20} + b_{0i}$ model the differences between the subjects at base-line. We shall therefore call them the cross-sectional components of the model. Further, the components $(\beta_{11} + b_{1i}) \ln(t)$ and $(\beta_{21} + b_{1i}) \ln(t)$ model how the responses evolve over time and are therefore called the longitudinal components in model (1). Obviously, the longitudinal component is of primary interest.

Model (1) is an example of a linear mixed model (Laird and Ware, 1982), which assumes that the outcome vector \mathbf{Y}_i for the i th subject satisfies

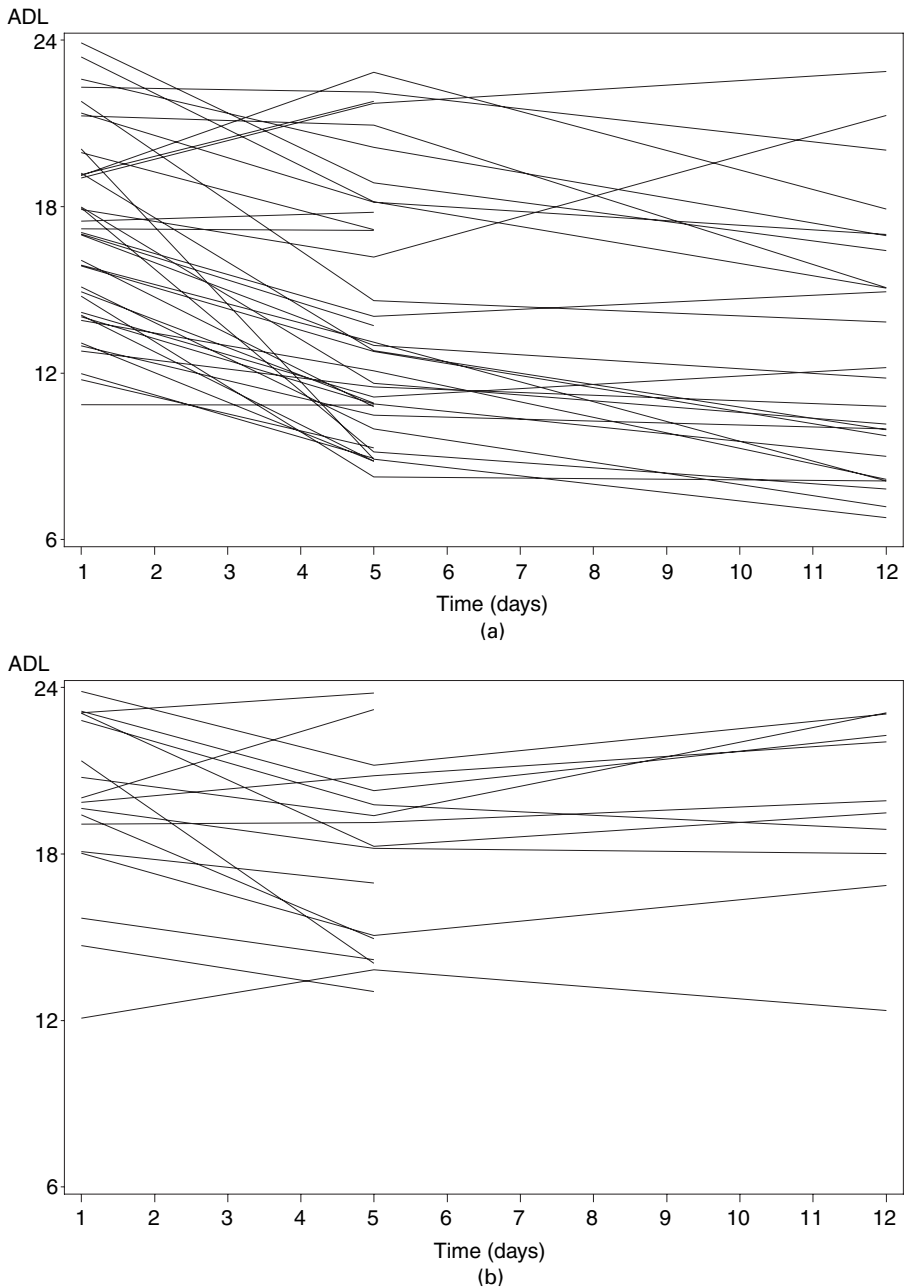


Fig. 1. ADL data: individual profiles for (a) non-neuropsychiatric and (b) neuropsychiatric patients

$$\mathbf{Y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad (2)$$

in which $\boldsymbol{\beta}$ is a vector of population-average regression coefficients called fixed effects, and where \mathbf{b}_i is a vector of subject-specific regression coefficients. The \mathbf{b}_i describe how the evolution of the i th subject deviates from the average evolution in the population. The matrices X_i and Z_i are $n_i \times p$ and $n_i \times q$ matrices of known covariates. The random effects \mathbf{b}_i and residual components

Table 1. ADL data: restricted maximum likelihood estimates, associated standard errors and Wald tests for fixed longitudinal effects, based on only those subjects for whom all three scheduled ADL measurements are available as well as based on all subjects†

Effect	Results for only 35 complete subjects		Results for all 54 subjects	
	Estimate	p-value	Estimate	p-value
<i>Linear mixed model without base-line covariates</i>				
Neuropsychiatric (β_{11})	-0.421 (0.454)	0.3544	-0.626 (0.376)	0.0962
Non-neuropsychiatric (β_{21})	-2.098 (0.287)	<0.0001	-1.926 (0.251)	<0.0001
Difference ($\beta_{11} - \beta_{21}$)	1.678 (0.537)	0.0018	1.300 (0.452)	0.0041
<i>Linear mixed model with base-line covariates</i>				
Neuropsychiatric (β_{11})	-0.421 (0.454)	0.3544	-0.589 (0.372)	0.1137
Non-neuropsychiatric (β_{21})	-2.098 (0.287)	<0.0001	-1.896 (0.248)	<0.0001
Difference ($\beta_{11} - \beta_{21}$)	1.678 (0.537)	0.0018	1.307 (0.447)	0.0035
<i>Conditional linear mixed model</i>				
Neuropsychiatric (β_{11})	-0.421 (0.454)	0.3544	-0.622 (0.376)	0.0981
Non-neuropsychiatric (β_{21})	-2.098 (0.287)	<0.0001	-1.920 (0.250)	<0.0001
Difference ($\beta_{11} - \beta_{21}$)	1.678 (0.537)	0.0018	1.298 (0.452)	0.0040
<i>Analysis of covariance</i>				
Neuropsychiatric (β_{11})	1.142 (0.782)	0.1534	1.146 (0.779)	0.1504
Non-neuropsychiatric (β_{21})	-2.010 (0.494)	0.0003	-2.043 (0.493)	0.0002
Difference ($\beta_{11} - \beta_{21}$)	3.153 (0.925)	0.0017	3.189 (0.922)	0.0015

†Four analyses are presented: a linear mixed model without base-line covariates, a linear mixed model with base-line covariates (age and living situation), a conditional linear mixed model and analysis of covariance. Standard errors are given in parentheses.

ε_i are assumed to be independent with distributions $N(\mathbf{0}, D)$ and $N(\mathbf{0}, \sigma^2 I_{n_i})$ respectively. Inference for linear mixed models is based on maximum likelihood or restricted maximum likelihood estimation under the marginal model for \mathbf{Y}_i , i.e. the multivariate normal model with mean $X_i\beta$, and covariance

$$V_i = Z_i D Z_i' + \sigma^2 I_{n_i}.$$

More specifically, we have that

$$\hat{\beta} = \left(\sum_i X_i' V_i^{-1} X_i \right)^{-1} \sum_i X_i' V_i^{-1} \mathbf{Y}_i, \quad (3)$$

which is normally distributed, with mean β and covariance $(\sum_i X_i' V_i^{-1} X_i)^{-1}$, from which inferences immediately follow. We refer to Verbeke and Molenberghs (2000) for more details on estimation and inference in the context of linear mixed models.

The estimation and inference results for the fixed slopes β_{11} and β_{21} , which are obtained from fitting model (1) to the ADL data, are shown in the top right-hand part of Table 1. We obtain a significant difference in the average evolutions over time ($p = 0.0041$), with no significant time trend for the neurocognitive patients and a significant average improvement for the patients who were not neurocognitive preoperatively. The variance of the random intercepts b_{0i} was estimated as 9.0968, which is large compared with the within-subject error variability σ^2 estimated as 2.4291. This suggests that there is much between-subject variability at base-line,

and therefore we wonder whether efficiency can be gained by formally correcting for these base-line differences. More specifically, it will be investigated how correction for base-line differences will affect estimation of and inference for the longitudinal component in a linear mixed model. In Section 3, we shall therefore discuss three different statistical procedures towards base-line correction.

3. Procedures for base-line correction

3.1. Introduction of base-line covariates

A classical statistical procedure that is often used to correct for differences between subjects is the inclusion of covariates which may explain the observed heterogeneity. In our ADL example, available potential covariates are the age of the patient at entry, as well as the living situation of the patient before the hip fracture. The living situation is a nominal variable indicating whether the patient was living on his or her own, living with a partner, family or friend(s), or living in a nursing home. Both variables are believed to be highly predictive of the functional status of the patients. In model (1), age and living situation can be formally corrected for by including them as covariates (after recoding the living situation by two dummy variables). Obviously, the advantage of this procedure is that the new model is still a linear mixed model such that the same estimation and inferential procedures can be used as for the original uncorrected model. Among the disadvantages is the fact that predictive covariates must be known and available, but also the fact that the correct parametric functional relationship between the covariates and the response of interest must be specified.

3.2. Conditional linear mixed models

An alternative way to make inferences about the longitudinal components in a linear mixed model, independently of the way that the cross-sectional components have been parameterized, is the use of conditional linear mixed models, which were introduced by Verbeke *et al.* (2001). The key idea is to rewrite model (1) as

$$Y_i(t) = \begin{cases} b_{0i}^* + (\beta_{11} + b_{1i}) \ln(t) + \varepsilon_i(t) & \text{if neuropsychiatric} \\ b_{0i}^* + (\beta_{21} + b_{1i}) \ln(t) + \varepsilon_i(t) & \text{if non-neuropsychiatric,} \end{cases} \quad (4)$$

with subject-specific intercepts b_{0i}^* , which are not of any interest. These nuisance parameters can be eliminated by conditioning on their sufficient statistics and inference for the longitudinal parameters of interest can be based on the likelihood of our data, conditional on the sufficient statistics for the nuisance parameters. A similar conditioning argument forms the basis of conditional logistic regression, which is a standard tool for the analysis of matched binary data (Breslow and Day (1989), section 7.1). In our context, this conditional inference can be shown to be equivalent with transforming all vectors $\mathbf{Y}_i \rightarrow A_i' \mathbf{Y}_i$ for any set of full rank $n_i \times (n_i - 1)$ matrices A_i satisfying $A_i' \mathbf{1}_{n_i} = 0$ and $A_i' A_i = I_{n_i - 1}$. Here, $\mathbf{1}_{n_i}$ is the n_i -dimensional vector containing only 1s, and $I_{n_i - 1}$ equals the $(n_i - 1)$ -dimensional identity matrix. The transformed data then satisfy a new linear mixed model with transformed covariates, and in which the only remaining parameters are the original longitudinal effects (fixed as well as random), and the residual variance σ^2 . Hence, standard software for linear mixed models can again be applied to obtain inferences for the parameters of interest, independently of any parametric assumptions about the base-line differences between the study participants. For two repeated measurements per subject, the transformation $\mathbf{Y}_i \rightarrow A_i' \mathbf{Y}_i$ reduces to the calculation of the within-subject differences between the two outcome measurements. Hence the conditional linear mixed model can

be viewed as an extension of the well-known paired t -test to longitudinal data sets, possibly unbalanced, with more than two measurements per subject.

Finally, it can be shown that the conditional linear mixed model yields the same estimates and inferences for the longitudinal effects as the linear mixed model (4) in which the subject-specific intercepts b_{0i}^* are considered unknown fixed parameters, and provided that restricted maximum likelihood estimation is used for parameter estimation (Verbeke and Molenberghs (2000), chapter 13). Hence, the conditional linear mixed model also has a fixed effects interpretation.

The theoretical properties of the estimates that are obtained from conditional linear mixed models have been investigated in detail by Verbeke *et al.* (2001). These properties immediately follow from the close relationship with restricted maximum likelihood estimation, for which theoretical justification can be found in Patterson and Thompson (1971), Sprott (1975) and Harville (1974, 1977).

3.3. Analysis of covariance

A third technique is analysis of covariance in which the first observed outcome is used as covariate in a model for the subsequent observations. For our ADL example, we would then consider the model

$$Y_i(t) = \begin{cases} \beta_{10} + b_{0i} + \gamma y_i(1) + (\beta_{11} + b_{1i}) \ln(t) + \varepsilon_i(t) & \text{if neuropsychiatric} \\ \beta_{20} + b_{0i} + \gamma y_i(1) + (\beta_{21} + b_{1i}) \ln(t) + \varepsilon_i(t) & \text{if non-neuropsychiatric,} \end{cases} \quad (5)$$

in which $y_i(1)$ is the first outcome value measured on subject i , i.e. the ADL score for subject i at day 1 post-operatively. Further, γ is a fixed unknown regression coefficient, and the model is restricted to outcomes $Y_i(t)$ measured at time points $t > 1$ which, in our example, is equivalent to $t = 5, 12$. Again, model (5) is a linear mixed model, implying that estimation and inference can be obtained from standard software.

In this particular example, model (5) is overspecified since the 2×2 covariance matrix for the second and third measurement is parameterized with four parameters, which is one too many. Any restriction can be used, e.g. assuming independence of random intercepts and slopes, or setting specific variance components equal to a prespecified value. All such analyses will lead to the same fitted normal model for \mathbf{Y}_i , with mean $X_i\beta$ and covariance matrix

$$V_i = Z_i D Z_i' + \sigma^2 I_{n_i}$$

and hence to exactly the same inferences for the elements in β .

4. Application: activities of daily living data

We shall now apply all three procedures to our ADL case-study. We hereby first focus on the 35 patients for which all three scheduled ADL measurements have been taken (Section 4.1). Afterwards, all methods will be applied to the complete data set with all 54 subjects (Section 4.2). All the results have been summarized in Table 1.

4.1. Balanced data

Balanced longitudinal measurements are obtained when a fixed number of measurements are taken for all subjects, at fixed time points. Often, models for balanced data have properties which are no longer valid in unbalanced situations (see for example classical analysis-of-variance models; Neter *et al.* (1990)). We therefore first compare the results of the various procedures obtained from the complete profiles, i.e. based on all subjects with all three scheduled measurements avail-

able. The results are shown on the left-hand side of Table 1. Note how the correction for the base-line covariates age and living situation have not affected the estimates for the longitudinal effects, nor their estimated standard errors. Still, the age of the patient as well as the preoperative living situation were found to be significantly related to the outcome of interest ($p=0.0032$ and $p=0.0477$ respectively; the results are not shown). Exactly the same estimates and standard errors are obtained from the conditional linear mixed model, which provides inferences for the longitudinal effects β_{11} and β_{21} , without making any parametric assumptions about the way that subjects differ at base-line. Finally, we observe that the results from the analysis of covariance are not consistent with the results from the other two methods. The estimates are distinctly different and the standard errors are larger.

4.2. Unbalanced data

We repeated the analyses after having added the 19 subjects for which only the first two measurements are available. The results are now shown on the right-hand side of Table 1. As is expected, we obtained different, more efficient results when compared with the results that are based on the complete profiles only. However, in contrast with before, we now also obtained (slight) differences in parameter estimates, as well as standard errors, between the first three procedures. As before, the analysis of covariance yields estimates that are distinctly different from those from the other procedures, as well as much larger standard errors.

5. Discussion

To explain the results that were obtained in Section 4, we first briefly summarize some theoretical results that were derived in Verbeke and Fieuws (2005), on the effect of including base-line covariates in linear mixed models, when interest is on inferences for longitudinal effects. We again consider the balanced case first, with exactly n observations per subject, all observations taken at fixed time points. It can then be shown that, in growth curve models, estimates and associated standard errors for longitudinal effects are not affected when base-line covariates are added to the model. Growth curve models (Laird *et al.*, 1987; Lange and Laird, 1989) are linear mixed models of the form (2), but with all design matrices X_i of the form $[\mathbf{1}_{n_i}|T] \otimes \mathbf{a}_i$ where $\mathbf{1}_{n_i}$ is the n -dimensional vector of 1s, T is an $n \times (r-1)$ design matrix modelling the slopes and \mathbf{a}_i is a vector with s subject characteristics that is used to explain variability between subjects in intercepts and slopes. Typically, the $r-1$ columns in T consist of hierarchical orders of polynomials of the time points at which the n measurements have been taken. However, more general models can be considered as well, e.g. piecewise linear models. The growth curve model basically assumes that the same covariates are used to model differences between subjects in the intercepts as well as differences between subjects in time effects. Furthermore, it is assumed that the design matrices Z_i are of the form $[\mathbf{1}_{n_i}|Z]$ for all i , where the columns in Z are the first $q-1$ columns of T , $q \leq r$. This automatically leads to a so-called well-formulated model (Morrell *et al.*, 1997) which does not include any polynomial effects or interactions unless all hierarchically lower order terms have been included as well.

Clearly, model (1) is a growth curve model, with T and Z equal to the vector containing the transformed time points $\ln(t)$, $t = 1, 5, 12$, and with \mathbf{a}_i equal to the vector with elements 1 and 0, or 0 and 1, depending on whether the subject belongs to the neuropsychiatric or non-neuropsychiatric group respectively. Hence, the above result explains why adding the base-line characteristics age and living situation does not affect the longitudinal conclusions. This would not hold for omitting the main effect of the neuropsychiatric status. Indeed, omitting this main

effect would no longer lead to a growth curve model as the interaction between neuropsychiatric status and $\ln(t)$ would still be in the model. Further, the fixed effects interpretation of the conditional linear mixed model approach shows that this model reduces to adding well-chosen base-line covariates to the original linear mixed model, again not affecting the results for the longitudinal effects. Finally, the above result implies that the longitudinal inferences that are obtained from the analysis of covariance would not be affected if the base-line response value $y_i(1)$ were left out of model (5). Hence, as far as the longitudinal trends are concerned, the analysis of covariance is equivalent to fitting the original linear mixed model (1) to the data restricted to the second and third measurements only, thereby completely ignoring the base-line measurements $y_i(1)$. This explains the observed increased standard errors for the estimates of β_{11} , β_{21} and $\beta_{11} - \beta_{21}$. Therefore, the only effect of analysis of covariance is loss of efficiency.

The analysis of covariance not only leads to much larger standard errors, but also to very different point estimates of the slopes. As a referee pointed out, one possible explanation might be a model violation. Indeed, if all three measurements are analysed, *versus* only the last two outcomes, then the linearity assumption will be very crucial. For the ADL data, however, we did not find any evidence for non-linearities ($p = 0.1114$ based on the complete profiles only; $p = 0.0844$ based on all profiles). This immediately points to an additional disadvantage of the analysis-of-covariance technique, i.e. the fact that, by restricting the analysis to the last two outcomes, information is lost about part of the evolution in the profiles.

In the case of unbalanced data, the results summarized earlier for growth curve models are no longer valid. In general, one can show that omitting base-line covariates will imply biased inferences for the longitudinal time trends of interest. This explains why the linear mixed models, with and without base-line covariates, and the conditional linear mixed model, no longer provided identical results for the longitudinal parameters of interest. Fortunately, the bias due to the omission of important base-line covariates can, in many applications, be expected to be small (Verbeke and Fieuws, 2005). In our example, this is probably reflected in the fact that, although not completely identical, the longitudinal inferences are very similar for the first three procedures. Finally, given our earlier remarks with respect to analysis of covariance for the balanced case, it is not surprising that strongly different estimates and much higher standard errors are obtained in the unbalanced case as well.

It should be emphasized that the theoretical results in Verbeke and Fieuws (2005), which completely explain our current findings, are not based on asymptotics arguments but solely use expression (3) for the estimate of the regression parameters in linear mixed models. Hence, our findings in the particular example of the ADL data are applicable in general, for small as well as large data sets. Also, although the small effect of base-line covariates on longitudinal inferences may not come as a complete surprise, unbalanced examples can be constructed where considerable bias would occur. These are situations where the omitted base-line covariates contain information about the nature of the imbalance in the data, e.g. about the time points at which measurements have been taken.

6. Conclusion

When analysing longitudinal data, primary interest is in studying how subjects evolve over time, and what subject characteristics affect these evolutions. Often, especially in observational studies, much base-line variability is observed between subjects. In this paper, we have addressed the question of whether or not we should correct for such base-line heterogeneity. Two procedures that are often used in practice to correct for base-line differences are the inclusion of important base-line covariates and the analysis of covariance. In the context of the ADL case-study,

we have shown that the inclusion of base-line covariates will have no, or very little, effect on the inferences for the longitudinal effects (the parameters of interest), and strong evidence has been found that analysis of covariance is very inefficient, especially in examples with a small number of repeated measurements per subject. We also have discussed and illustrated the use of conditional linear mixed models which provide inferences for the longitudinal component of the model, completely independently of any parametric assumption about the way that subjects differ at base-line.

The key argument in explaining our findings was the fact that longitudinal inferences based on linear mixed models are very robust with respect to the inclusion of base-line covariates. This is no longer expected to be so for non-linear mixed models (Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1997) or generalized linear mixed models (Pinheiro and Bates, 2000). It therefore should be emphasized that our findings apply only to the context of linear mixed models for continuous longitudinal data. Future research is required to study how our results extend to other types of models or data.

References

- Bonate, P. L. (2000) *Analysis of Pretest-posttest Designs*. Boca Raton: Chapman and Hall–CRC.
- Breslow, N. E. and Day, N. E. (1989) *Statistical Methods in Cancer Research*, vol. 1, *The Analysis of Case–Control Studies*. Lyon: International Agency for Research on Cancer.
- Davidian, M. and Giltinan, D. M. (1995) *Nonlinear Models for Repeated Measurement Data*. New York: Chapman and Hall.
- Delesie, L., Sermeus, W. and Vanden Boer, G. (1987) Syllabus for the formation with regard to medical activities, kings decision 14/08/87, activities of daily living. *Technical Report*. Ministry of Public Health and Ministry of Social Affairs, Brussels, and Centre for Health Services Research, Katholieke Universiteit Leuven, Leuven.
- Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994) *Analysis of Longitudinal Data*. Oxford: Clarendon.
- Harville, D. A. (1974) Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**, 383–385.
- Harville, D. A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Statist. Ass.*, **72**, 320–340.
- Katz, S. and Akpom, C. A. (1976) A measure of primary sociobiological functions. *Int. J. Hlth Serv.*, **6**, 493–507.
- Laird, N. M., Lange, N. and Stram, D. (1987) Maximum likelihood computations with repeated measures: application of the EM algorithm. *J. Am. Statist. Ass.*, **82**, 97–105.
- Laird, N. M. and Ware, J. H. (1982) Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.
- Lange, N. and Laird, N. M. (1989) The effect of covariance structure on variance estimation in balanced growth-curve models with random parameters. *J. Am. Statist. Ass.*, **84**, 241–247.
- Milisen, K. (1999) An intervention study for delirium in elderly hip fracture patients. *PhD Thesis*. Katholieke Universiteit Leuven, Leuven.
- Milisen, K., Abraham, I. L. and Broos, P. L. O. (1998) Postoperative variation in neurocognitive and functional status in elderly hip fracture patients. *J. Adv. Nursng*, **27**, 59–67.
- Morrell, C. H., Pearson, J. D. and Brant, L. J. (1997) Linear transformations of linear mixed-effects models. *Am. Statistn*, **51**, 338–343.
- Neter, J., Wasserman, W. and Kutner, M. H. (1990) *Applied Linear Statistical Models: Regression, Analysis of Variance and Experimental Designs*, 3rd edn. Boston: Irwin.
- Patterson, H. D. and Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–554.
- Pinheiro, J. C. and Bates, D. M. (2000) *Mixed Effects Models in S and S-Plus*. New York: Springer.
- Sprott, D. A. (1975) Marginal and conditional sufficiency. *Biometrika*, **62**, 599–605.
- Verbeke, G. and Fieuws, S. (2005) The effect of misspecified baseline characteristics on inference for longitudinal trends in linear mixed models. To be published.
- Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Verbeke, G., Spiessens, B. and Lesaffre, E. (2001) Conditional linear mixed models. *Am. Statistn*, **55**, 25–34.
- Vonesh, E. F. and Chinchilli, V. M. (1997) *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Dekker.