original article

Analysing longitudinal data with multilevel modelling

Nikos Ntoumanis
Curtin University

Recent advances in statistical analysis have resulted in the decline of the use of repeated measures ANOVA/MANOVA for the analysis of longitudinal data in health psychology research. One of the more sophisticated and comprehensive alternatives to these tests (see Kwok et al., 2008) is multilevel modelling (MLM), also known as hierarchical linear modelling or linear mixed modelling. MLM is appropriate for the analysis of data with a nested structure, for example, patients (level 1) nested within clinics (level 2). Ignoring the nested structure of such data can result in biased estimates of standard errors and subsequent increase in Type I error (Hox, 2010). MLM is also useful for testing the interaction between individual and contextual factors and exploring heterogeneity in the data due to their nested structure. Many applications of MLM in the health psychology literature incorporate two or three levels of analysis. For example, Mayberry, Espelage and Koenig (2009) examined adolescents' perceptions of parental and peer influence (level 1) and school characteristics (level 2) as predictors of adolescent substance use. In addition to analysing cross-sectional data, MLM can also be used for longitudinal data, given that multiple measurement points (level 1) are nested within individuals (level 2) who can also be nested within a group setting (level 3). For example, Ntoumanis, Taylor and Thogersen-Ntoumani (2012) examined moral attitudes, emotional well-being, and indices of behavioural investment in a sample of British adolescent athletes. In this study, each variable was measured at three time points during a sport season. The three time points were the first level of the analysis, with athletes and their teams constituting the second and third levels of the analysis, respectively.

In this paper, I offer a very brief overview of how multilevel modelling can be employed for the analysis of longitudinal data without presenting any mathematical formulas. I use an example from the physical activity literature to demonstrate, step-by-step, decisions that need to be made with regard to the analysis of the data. I refer the reader to Singer and Willett's (2003) book for a far more detailed treatment of MLM for longitudinal data analysis, including testing the assumptions that underlie such analysis.

MLM can be used when all individuals are assessed on the same number of occasions which are equally spaced over time. However, MLM can also be used when the spacing of measurement points is not identical across individuals (e.g., the time interval between cancer screenings might vary across participants), and also when the number of measurement waves is not the same across individuals. The latter is a particularly important feature, given the attrition of participants recorded in longitudinal studies. As Singer and Willett (2003) note, each individual's growth record can contain a unique number of waves collected at unique occasions of measurement. The impact of missing data on MLM estimates is discussed by Hox (2010).

In its simplest form, a MLM of change is a linear growth model with a random intercept, as well as a random slope to represent change over time in the dependent variable (note that a model with no growth term can also be calculated initially in order to estimate the intraclass-correlation coefficient which quantifies the variation in the dependent variable across the different levels of the analysis). For example, Figure 1 demonstrates changes in intrinsic motivation (measured on a 1-7 point scale with higher scores indicating greater motivation) for physical activity in children over 6 time points in a three-year period. It is clear from the figure that there was considerable inter-individual variation in the motivation scores at the beginning of the study and in the trajectory of change of motivation over time. Such heterogeneity in the intercept and the rate of change can be captured by including additional predictors in the model. The intercept and growth across the whole sample are shown with the thick dotted line. Such variations cannot be captured in a fixed effects ANOVA model, but can be important from an applied and conceptual perspective. Multilevel modelling provides a statistical test of the variation in both the intercept and the growth terms across individuals (see Model 1 in Table 1).

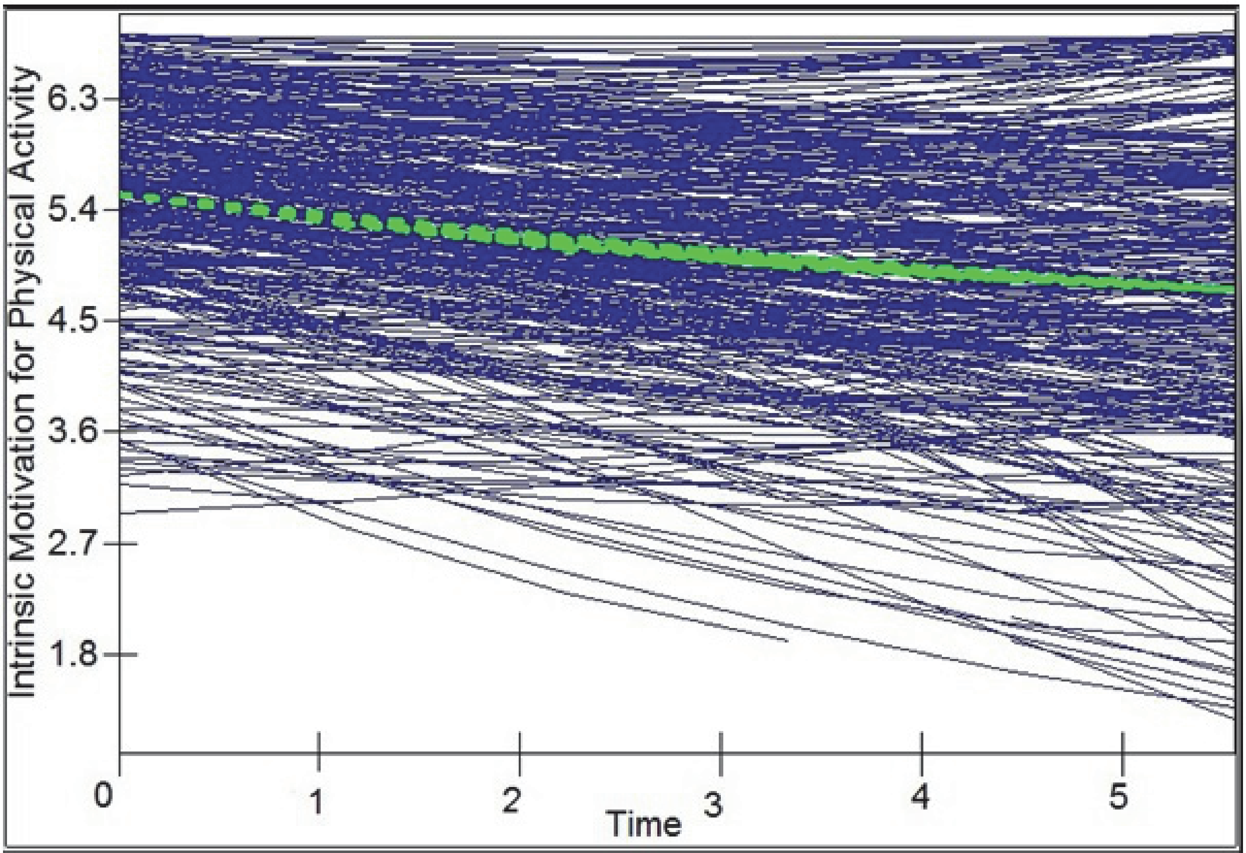

Figure 1: Variation in intercepts and slopes of change in intrinsic motivation for physical activity over six time points

One of the first issues to think about when using MLM to analyse longitudinal data is what type of change to examine. In our example, up to three growth terms can be tested: linear, quadratic and cubic. A study will need at least 3, 4 and 5 time points to test interindividual variation in linear, quadratic, and cubic growth, respectively. Figure 1 includes both linear ($b = -.25$, $p < .01$) and quadratic ($b = .02$, $p = .03$)

Table 1. *Model Examining Changes in Intrinsic Motivation for Physical Activity Over a 3-Year Period*

	Model 1 *B (SE)*	Model 2 *B (SE)*	Model 2 *B (SE)*
Level 1 Predictors			
Intercept	5.526 (.066)	5.529 (0.60)	4.292 (3.15)
Time (linear)	-.254 (.046)	-.249 (.043)	-.249 (.043)
Time (quadratic)	.020 (.009)	.020 (.008)	.020 (.008)
Competence		.320 (.051)	.234 (.055)
Competence x Time (linear)		.071 (.040)	.090 (.040)
Competence x Time (quadratic)		-.008 (0.008)	-.011 (.008)
Level 2 Predictors			
Mean Competence			.249 (.062)
Level 1 Variance	.656 (.033)	.524 (.028)	.521 (.028)
Level 2 Variance			
Intercept	.945 (.119)	.703 (.100)	.694 (.099)
Time (linear)	.155 (.062)	.153 (.055)	.156 (.055)
Time (quadratic)	.004 (.002)	.005 (.002)	.005 (.002)
Competence		.116 (.026)	.112 (.026)
Model Deviance	5507.596	5021.247	5185.553

Note: When the ratio of B/SE is above 1.96, the parameter is considered significant at $p < .05$; this ratio is a z statistic

growth; the cubic growth term was almost zero and was excluded from the equation. Both linear and quadratic terms were significant and their inter-individual variation was also significant (see Model 1 in Table 1). Although not shown in Table 1, it is also useful to inspect covariances between the intercept and the growth terms in order to determine whether participants' initial mean score of intrinsic motivation is related to the rate of change of their scores over time. However, note that it is not always necessary that the intercept represents initial status. In multilevel growth models the growth term or terms for time can be centred by assigning the value of zero to different time points, such as the beginning, middle or end of the study (or any time point of interest), depending on the substantive question pursued in the study. In Figure 1, time is centred (time1 = 0) at the first wave of measurement, hence the intercept of the growth model can be interpreted as students' reports of intrinsic motivation at beginning of the study.

Another issue to consider when analysing longitudinal data with MLM is the type of predictors that can be included in the analysis. In addition to the intercept and the growth terms, additional predictor variables can be added in the multilevel regression equation at different levels of the analysis. In a two-level model (repeated measures nested within individuals) both time-varying covariates (level 1) and time-invariant covariates (level 2) can be introduced. By adding predictors the unaccounted variance at the corresponding level of each predictor can be reduced, however, the unaccounted variance at the other level might either decrease or increase. Singer and Willett (2003) discuss this problem and suggest suitable pseudo-R2 indices.

An example of a level 1 covariate in our example could be perceptions of physical activity competence measured across all six time points. An example of a level 2 covariate could be a personality or demographic variable measured at one point in time. Furthermore, interactions can be tested between predictor variables within the same level or at different levels. In our example, we entered in the multilevel regression equation the additional predictors of perceptions of physical activity competence, as well as the interactions between competence and linear growth and between competence and quadratic growth (see Model 2 in Table 1). Perceptions of competence emerged as significant predictors of intrinsic motivation ($b = .32$, $p < 0.01$) but the two interaction terms were not significant. The effect of each level 2 predictor can be tested as fixed or random. In our example, the main effect of competence could, depending on available theory or evidence, be conceptualised as being the same across all individuals and therefore treated as fixed, or varying from individual to individual and hence treated as random. Whilst treating the slopes of level 2 predictors as random helps researchers answer interesting research questions associated with between-person variability, a model with many random effects might not converge. Singer and Willett (2003) and Hox (2010) offer some detailed guidance for model building and model comparison, involving the inspection of deviance statistics for each model.

In growth models the slope of a level 1 (time-varying) predictor confounds inter-individual change and between-person variability. Hence, it is suggested that the aggregate of each level 1 predictor is entered at the level 2 of the analysis. In our example, if we average perceptions of competence within each individual across all measurement waves, this variable could be entered as a level 2 predictor in the analysis. In the new model (Model 3, Table 1), the slope of the level 1 measure of competence ($b = .23$, $p < 0.01$) represents the within-person association between competence and intrinsic motivation over time, after partialling out between-person differences in competence. However, the two slope terms for competence at the two levels of the analysis might or might not be correlated, depending on how the level 1 predictor has been centred. The issue of centring is often discussed in the MLM literature and is another important factor to consider with this type of analysis. Centring helps the interpretation of their intercepts and slopes but the type of centring has often puzzled researchers unfamiliar with the complexities of the analysis. Enders and Tofighi (2007) and Lüdtke, Robitzsch, Trautwein and Kunter (2009) offer some excellent guidance on centring for cross-sectional multilevel data, and their recommendations also apply for longitudinal MLM data.

Briefly stated, the level 1 predictor scores could be centred around each person's unique mean score over time (group-mean centring; CWC) or across all individuals' mean score over time (grand-mean centring; CGM). In both cases, the level 1 slope is the same, but the level 2 slope will differ. With CGM, the level 1 and 2 slopes are correlated, hence the level 2 slope is a partial effect controlling for level 1. With CWC, the two slopes are uncorrelated, hence the level 2 effect is a mixture of level 1 and level 2 effects (this is the case for the level 2 slope for competence shown in Model 3, Table 1). To obtain a pure estimate of level 2 effect, we need to calculate the difference between the level 2 and level 1 slopes (Marsh et al., 2012); in our example, .249- .234= .015. In brief, if the within-person associations (level 1) are of interest, then either type of centring will provide the same estimate which is not confounded by inter-individual differences. However, if inter-individual differences are of interest, then the type of centring used will result in different slope estimates. In most cases, however, researchers are interested in Level 1 associations.

Testing linear or non-linear terms for time in a MLM equation is a sensible option when certain trends are expected over time. In other studies (e.g., diary studies) such trends might not be expected. For example, if one is interested in examining dietary lapses over a typical 7-day period, there is no rationale to expect a particular pattern of growth over that period. However, other contrasts of interest could be entered to detect specific trends. For example, McKee, Ntoumanis and Taylor (in press) showed that dietary lapse occurrences were more likely in the evening compared to the morning ($b=0.37$, $p=0.002$) and afternoon ($b=0.24$, $p=0.01$) over a 7-day period.

Often in health psychology researchers are interested in several dependent variables. In such cases a multivariate growth model can be used instead of several univariate growth models. Specialised MLM software such as MLwiN can perform this analysis by adding one extra level. Other software with structural equation modelling capabilities (e.g., Mplus) can also perform multivariate MLM but with a different set up; in fact, in Mplus the number of levels is one less than the number of levels in conventional MLM software (Muthén & Muthén, 1998-2012). Such software can also perform multilevel structural equation modelling which, unlike standard applications of MLM regressions, take into account measurement error and can test both simple and complex mediation effects (Preacher, Zyphur, & Zhang, 2010).

An often asked question revolves around the sample size needed to perform MLM analysis of change. Various rules of thumb have been proposed in the literature; for example, a simulation study by Maas and Hox (2008) suggests that sample sizes of 50 or less at level 2 can result in biased estimates of the standard error of the variance terms in that level. The regression coefficients and level 1 variance terms are fortunately not affected by this bias. Maas and Hox's simulation had 5 observations as the minimum number at level 1 (in other words, number of repeated measures for each individual in a longitudinal MLM). A better option than rules of thumb and simulation studies is the use of specialised software to calculate the sample size requirements for a particular study. A freely available software for power analysis, for both cross-sectional and longitudinal MLM is Optimal Design, available at http://sitemaker.umich.edu/group-based/optimal_design_software . For a two-level longitudinal MLM, the software requires input of values regarding the duration of the study, the frequency of observations, the level 1 variance, the between-person variability in the parameter of interest, and an estimate of effect size. It is also important that researchers build in estimates of expected attrition rates in their calculations.

In sum, MLM can address all the research questions that repeated measures ANOVA/MANOVA tests address without being constrained by the rigid assumptions of the latter (see Kwok et al., 2008). Further, MLM can be used to pursue research questions that cannot be answered with repeated measures ANOVA/MANOVA. Health psychologists can benefit in many ways from using MLM in their analysis of longitudinal data. Many commercial (etc., MLwiN, HLM, Mplus, SPSS, SAS) and some free software (R) can be used for such analysis.

References

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue.

Psychological Methods, 12(2), 121-138. doi:10.1037/1082-989X.12.2.121

Hox, J. J., (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

Kwok, O. M., Underhill, A. T., Berry, J. W., Luo, W., Elliott, T. R., & Yoon, M. (2008). Analyzing longitudinal data with multilevel models: An example with individuals living with lower extremity intra-articular fractures. *Rehabilitation Psychology, 53*(3), 370-386. doi:10.1037/a0012765

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*(2), 120-131. doi:10.1016/j.cedpsych.2008.12.001

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*(3), 86-92. doi:10.1027/1614-2241.1.3.86

McKee, H., Ntoumanis, N., Taylor, I.M. (in press). An ecological momentary assessment of lapse occurrence in dieters. *Annals of Behavioral Medicine*. doi: 10.1007/s12160-014-9594-y

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47*(2), 106-124. doi:10.1080/00461520.2012.670488

Mayberry, M. L., Espelage, D. L., & Koenig, B. (2009). Multilevel modeling of direct effects and interactions of peers, parents, school, and community influences on adolescent substance use. *Journal of Youth and Adolescence, 38*(8), 1038-1049. doi:10.1007/s10964-009-9425-9

Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Ntoumanis, N., Taylor, I. M., & Thøgersen-Ntoumani, C. (2012). A longitudinal examination of coach and peer motivational climates in youth sport: Implications for moral attitudes, well being, and behavioral investment. *Developmental Psychology, 48*(1), 213. doi:10.1037/a0024934

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*(3), 209-233. doi:10.1037/a0020141

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Nikos Ntoumanis
is Professor of Psychology at the Centre of Health Psychology and Behavioural Medicine, School of Psychology and Speech Pathology, Curtin University, Australia

nntoumanis@gmail.com