

Hypothesized, Directly-Coded Curve Shapes in Growth Curve Analysis: An Example

Patricia M. Herman
RAND Corporation

Lee Sechrest
University of Arizona

Growth curve analysis provides important informational benefits regarding intervention outcomes over time. Rarely, however, should outcome trajectories be assumed to be linear. Instead, both the shape and the slope of the growth curve can be estimated. Non-linear growth curves are usually modeled by including either higher-order time variables or orthogonal polynomial contrast codes. Each has limitations (multicollinearity with the first, a lack of coefficient interpretability with the second, and a loss of degrees of freedom with both) and neither encourages direct testing of alternative hypothesized curve shapes. Especially in studies with relatively small samples it is likely to be useful to preserve as much information as possible at the individual level. This article presents a step-by-step example of the use and testing of hypothesized curve shapes in the estimation of growth curves using hierarchical linear modeling for a small intervention study.

Keywords: growth curve analysis, non-linearity, hierarchical linear modeling, intervention outcomes

It is often the case that the success of an intervention is judged by its outcomes either at the end of the intervention or at some follow-up time period. Although that approach focuses the determination of effectiveness on a key point in time for the intervention under study, it provides no information on how patients¹ arrived at that point. For example, did patients improve at a constant rate over the entire period, did they achieve their end point levels quickly and then remain there for the rest of the study period, or did the impact of the intervention need to accumulate for most of the treatment period before any change in outcomes was seen? If the study has a follow-up period, did patients remain at their end-of-treatment levels, did they go on to improve beyond them, or did they begin to regress? Such information is essential to the determination of optimal treatment frequency and duration for practice and future research.

If outcome measures are available over 3 or more time points, a growth curve analysis is the logical approach to answer these questions. Instead of comparing only end points between groups, the “growth curves” (the trajectory of change in outcomes over time) of participants in each group can be compared over the full-study data collection period.

¹ In the context of the study to be presented, “patients” is an appropriate term (Kihlstrom, 2005), but the methodology to be described is relevant to subjects of any sort when studied over time.

Growth curve analyses can be performed as a special case of hierarchical linear modeling (HLM, also known as multilevel models, random-coefficients models, random-effects models, mixed models, and seemingly unrelated regression (Heo et al., 2003; Raudenbush & Bryk, 2002; Singer, 1998)). HLM is appropriate for growth curve analyses because there are at least two hierarchical levels of analysis: within participant (changes across time for each participant) and across participants (changes to each participant's growth curve intercept and slopes due to covariates). Growth curve analysis via HLM offers a number of statistical benefits in addition to the informational benefits mentioned above. For example, it corrects for the error structure violations that would be caused by performing an analysis of all data points using ordinary least squares (OLS) regression (i.e., observations over time for one individual cannot be assumed to be independent) (Feldman, 1988; Heo et al., 2003). HLM also allows for unbalanced designs (having a different number of observations in each group is troublesome to methods such as repeated measures ANOVA (O'Connell & McCoach, 2004) and two-step OLS methods (Heo et al., 2003)), and because it uses all information available in the dataset, it optimizes estimation in the presence of missing data (O'Connell & McCoach, 2004; Raudenbush & Bryk, 2002). This last benefit is especially important as missing data are common in studies in which repeated measurements are made over time on each subject.

If growth curve analysis is the method of choice, it still remains to decide how to model the shape of the curve over the intervention period and any change in shape during follow-up. Because higher-order (e.g., quadratic, cubic) time variables are highly correlated with each other and with the linear time variable, orthogonal polynomial contrast coding is often used to capture non-linearity in individual growth curves (see for example (Kurdek, 1999)). However, Biesanz et al. argue against orthogonal polynomial contrasts in favor of the use of "regular" polynomial coding on the basis of interpretability (Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004). In either case, if the number of data points available over time is limited (and especially if missing data are a problem), the degrees of freedom cost of the additional time variables may restrict the curve shapes possible.

Finally, neither HLM nor polynomial coding promotes the hypothesis and testing of theory-derived curve shapes. The use of theory in evaluation not only guides and strengthens research design, it also provides a context for the interpretation of results (Lipsey, 1990). Although both statistical approaches allow testing whether the curve has one or more "bends," neither specifies the alternatives considered theoretically feasible, thus giving all shapes equal credence.

The study reported here uses data from a randomized controlled trial for back pain to present an example of growth curve analysis using a third

alternative: direct comparison of hypothesized, directly-coded growth curve shapes. This approach avoids multicollinearity between time variables, maintains coefficient interpretability, minimizes the loss of degrees of freedom, and formalizes the use of theory in the determination of growth curve shape.

Method

Sample

The data used in this example are from a 2005 worksite study of alternative care for low back pain. This dataset has a number of qualities that make it a good example for this approach. It is a small sample (70 participants in total with 39 randomized to the treatment group and 31 to control) with up to five data points for each participant, and these five data points cover two periods expected to have different outcome trajectories. The first four cover monthly measurements from baseline to end of the 3-month treatment period, and the last is the end point of a 3-month follow-up period. In addition, as an incentive for retention, each participant in the control group was offered the opportunity to receive one month of treatment at the end of the intervention period. There were 13 participants in this “crossover” group. Finally, there are missing data, assumed to be missing at random. A complete dataset would contain 350 data points (70 times five). This dataset contains 314 data points for the 70 participants. Most missing data occurred for the month 3 and month 6 data points

Outcomes

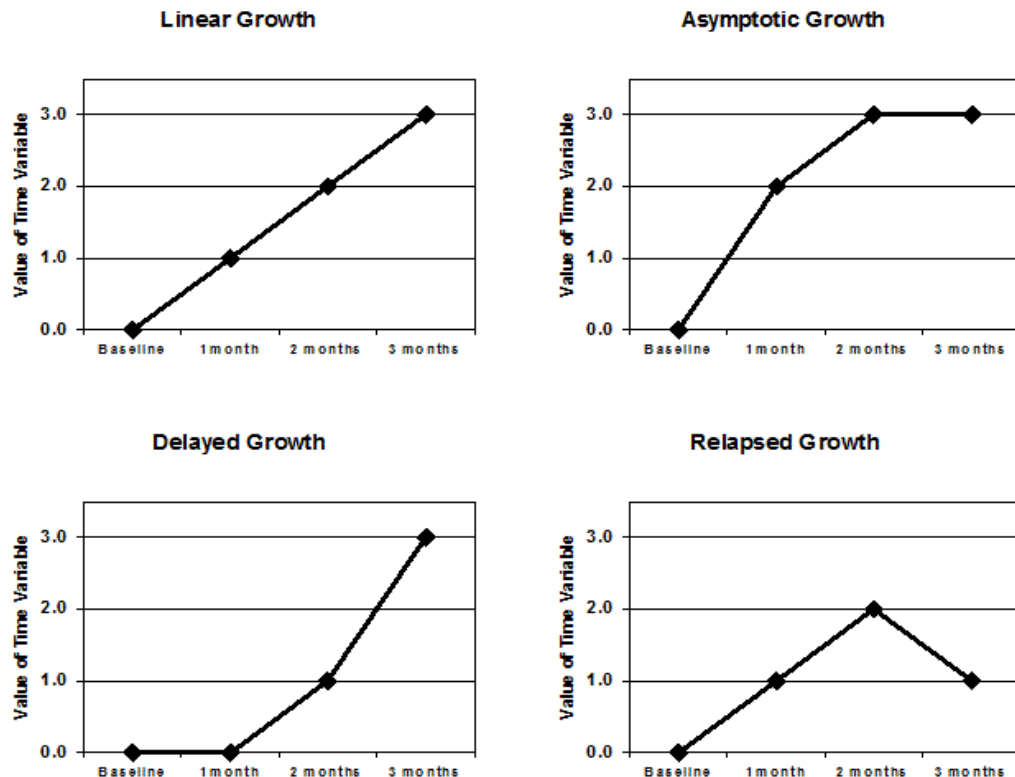
Three outcomes were measured in this study: the Roland Disability Questionnaire (RDQ) (Roland & Morris, 1983), the Oswestry Disability Index (ODI) (Fairbank, Couper, Davies, & O'Brien, 1980), and a pain visual analog scale (VAS) (Huskisson, 1982). More detail on the original study and its intervention and results can be found in Szczurko et al. 2007. The main outcome used in this example is the RDQ. Robustness checks of the RDQ growth curve model estimates are performed using the ODI and the pain VAS.

Statistical Analysis

Four basic shapes (shown in Figure 1) are hypothesized as plausible to represent the dominant growth curve shape experienced by participants during the 3-month intervention period. These four shapes are based on common healing trajectories seen in practice: steady linear growth (steady change—i.e., either constant improvement or worsening of symptoms over

time), asymptotic growth (rapid change tapering off over time as some ceiling or floor is reached), delayed growth (slow change at first, then rapid change—possibly due to the need for the intervention effects to accumulate to some threshold before change is seen), and relapsed growth (rapid change followed by a return toward baseline levels—often seen in intractable conditions where hope or novelty will cause a temporary, but unsustainable change in reported outcomes). The curves in Figure 1 are all drawn showing an initial increase from baseline. Since a reduction from baseline constitutes an improvement in the outcomes measured in this study, a negative coefficient is expected for the shape (time) variable for both groups. It is important to note that other possible growth trajectories were excluded as not plausible (assuming effectiveness of treatment). For example, we did not expect to find an inverted form of the relapse function, i.e., getting worse initially, followed by improvement (the “bottoming out” function). We did not expect to find delayed worsening. In this study, we were looking for the likelihood of specific expected trajectories of change.

Figure 1. Hypothesized growth curve shapes during the 3-month intervention period



HYPOTHESIZED CURVE SHAPES IN GROWTH CURVE ANALYSIS

The hypothesized shapes are coded into four versions of the time variable as shown in Table 1. The baseline value for time in each case is coded as zero to allow for the intercept to have a meaningful interpretation—as the grand mean of baseline outcome values (Biesanz et al., 2004; Mehta & West, 2000; Singer, 1998). For the same reason all explanatory variables (other than group designation) are centered at their grand mean values. Because of the multitude of different combinations possible between the shape of the study period curve and the follow-up period curve (e.g., return to baseline, remain constant at the end-of-intervention value, continue along the same trend as shown during the intervention period, etc.), a separate time variable is used for the 3-month follow-up period are shown in Table 1 (Raudenbush & Bryk, 2002; West, Wu, & Ryu, 2006). We expected that participant outcomes for the treatment and non-crossover control group would remain at their 3-month end-of-intervention values during the follow-up period—thus, 3-month and 6-month values for the time variable for each shape are the same. The significance of the follow-up time variable will test that hypothesis.

Table 1
Coding for the Four Hypothesized Shapes for the Intervention Period Time Variable and for the Follow-up Period Time Variable

Shape	Baseline	1 Month	2 Months	3 Months	6 Months
Linear	0	1	2	3	3
Asymptotic	0	2	3	3	3
Delayed	0	0	1	3	3
Relapsed	0	1	2	1	1
Follow-up	0	0	0	0	1

The models are estimated using hierarchical linear modeling (HLM) allowing the intercept and two slopes (one for each time variable) to have both fixed (group mean) and random (participant-specific) effects. The unconditional models (models without inter-individual predictor variables) are first estimated for each of the four possible time (curve shape) variables. Unconditional models estimate the grand mean intercept and grand mean intervention and follow-up period slopes for the whole sample, as well as the variance and covariance across participants of these estimates. A growth curve analysis typically begins with the fitting of an unconditional model as it provides baseline statistics and useful empirical evidence for proper model specification (Raudenbush & Bryk, 2002). The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) fit statistics are compared across models to determine which growth curve shape best fit the sample as a whole and each group. Both fit statistics are based on the log likelihood, are in the form of “smaller is better,” and penalize the log likelihood for the number of parameters

estimated (Raudenbush & Bryk, 2002; Singer, 1998). If the difference between BIC statistics for two models is greater than 10, that can be considered “very strong evidence” that the model with the lower BIC is a better fit to the data (Raftery, 1995).

Once the best growth curve shape was determined, the variables indicating the treatment group status and the portion of the control group who elected to receive treatment after the intervention period ended (the crossover group) were added. This conditional model allows treatment and crossover group status to affect the intercept and both time variable slopes. After interpretation of these results, participant characteristic covariates were added as needed. Available participant characteristic variables were age, sex, whether the participant entered the study with comorbid conditions (yes=1; no=0), job position and shift (day, afternoon, or night shift warehouse worker, delivery person, truck driver, or sales representative with day shift as the reference), and the total number per week of visits to other practitioners for adjunctive back pain care during each period. Nested models were compared using a χ^2 test calculated as -2 times the difference between the log likelihood of the nested model and the log likelihood of the more inclusive model (Raudenbush & Bryk, 2002; Singer, 1998). The degrees of freedom for this test are equal to the number of variables in the more inclusive model minus the number in the nested (restricted) model. The results of the chosen (“best”) growth curve model were compared to the results of an end-point analysis focusing only on 3-month (end of the intervention period) outcomes as an illustration of what can be gained by the estimation of growth curves.

As a measure of the robustness of the estimated model, parameter estimates were applied to the ODI and pain VAS data (standardized to the RDQ mean and standard deviation) and residuals were checked for bias (the average value for residuals should be zero), and overall model fit (in terms of percentage of variance explained—regression sum of squares divided by total sum of squares) was compared. It was expected that since the pain VAS is known to correlate more highly with the RDQ than the ODI (Beurskens, de Vet, & Koke, 1996), the application of the estimated model parameters to the pain VAS would result in a fit closer to that of the RDQ than will be found for the ODI.

The model was estimated using restricted (residual) maximum likelihood (REML) estimation in the PROC MIXED procedure of SAS for Windows, Version 9.1 (SAS Institute Inc., Cary, NC). Coding for growth curve analysis was based on Singer (1998).

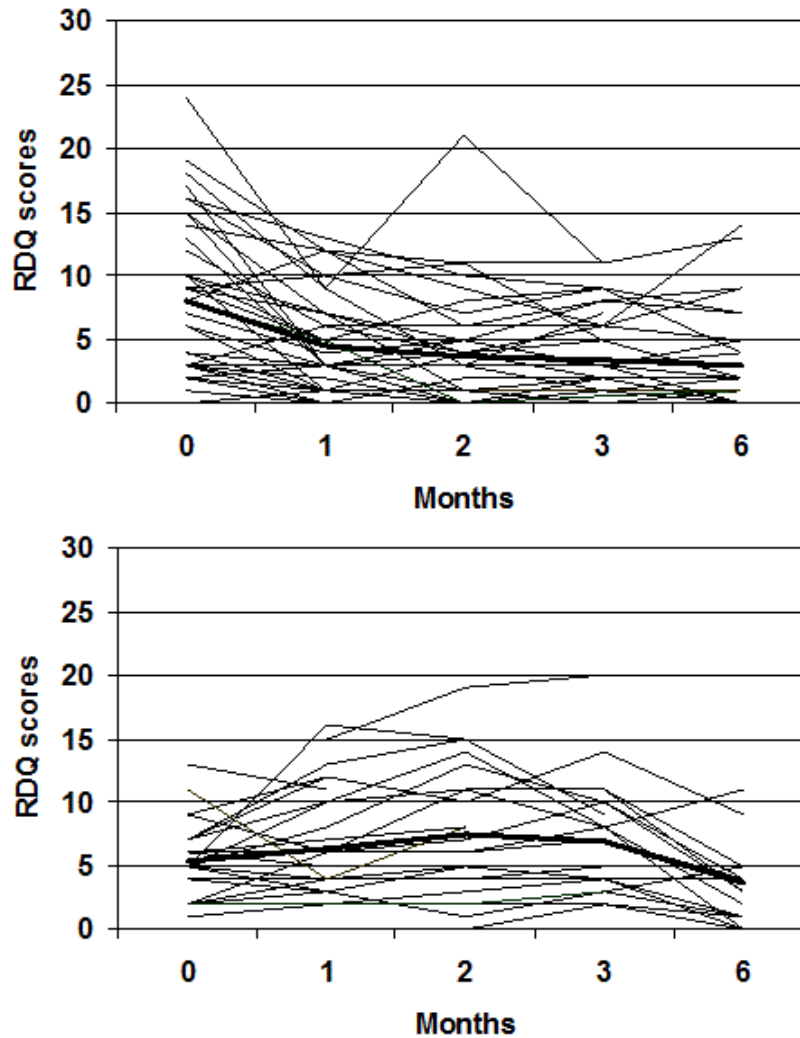
Results

Figure 2 contains parallel plot (spaghetti plot) graphs of each participant’s actual change in RDQ scores over the 3-month intervention

HYPOTHESIZED CURVE SHAPES IN GROWTH CURVE ANALYSIS

period and for the 3-month follow-up. The thicker black line in each graph indicates the average RDQ score for that group's participants over time. As can be seen, the average shape for the treatment group during the intervention period seems to be asymptotic, but the shape is indeterminate for the control group.

Figure 2. Lines drawn between the raw Roland-Morris Disability Questionnaire (RDQ) scores for each member of the treatment group (top) and each member of the control group (bottom). The darker line is the plot of the average RDQ scores over time for each group. The crossover group is included with the rest of the control group in the computation of the average RDQ scores in the bottom graph.



Determination of the Best Curve Shape

The results of the unconditional models for each hypothesized intervention period growth curve shape indicate that the asymptotic curve results in the best fit for the sample as a whole and for each group. The AICs and BICs for each shape for all three unconditional models are shown in Table 2. The asymptotic growth curve shape results in the smallest AIC and smallest BIC for all participants and for both groups. The difference between the asymptotic model BIC and the next smallest (the relapsed shape for all participants and the control group, and the linear shape for the treatment group) indicates strong evidence that the asymptotic shape gives the best fit to the data for all participants and for the treatment group, and indicates positive evidence, if not so strong, for the control group.

Table 2

Unconditional Model Fit Indices for Each Hypothesized Shape

	Linear growth	Asymptotic growth	Delayed growth	Relapsed growth
All participants – AIC	1667.8	1634.6	1693.1	1669.8
All participants – BIC	1683.5	1650.3	1706.6	1683.3
Treatment group only – AIC	965.8	932.1	1001.2	968.6
Treatment group only – BIC	975.8	942.1	1009.6	978.6
Control group only – AIC	639.8	635.7	648.7	638.9
Control group only – BIC	649.9	645.7	657.4	648.9

Note. AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion

Unconditional Model Results

The complete results of the asymptotic unconditional model are shown in Table 3. As can be seen, the average intercept (baseline value of the RDQ) across the sample is 6.79. On average, participants improve (go down) from their baseline values by 0.56 RDQ points per month. Of course, since time is modeled as an asymptotic curve, participants on average improve from baseline by 2×0.56 (or 1.12 RDQ points) in the first month, 3×0.56 (or 1.78 RDQ points) in the second month, and then remain at that value to the end of the intervention period. During the follow-up period participants on average improved significantly (went down) from their end of intervention (3-month) values by another 1.28 RDQ points.

The variance, covariance, and residual values at the bottom of the table provide valuable baseline information on the sources of variation in individual RDQ values across participants. That there is significant variance across participants' initial status (intercept) and rate of change during the intervention period (time slope) justifies a model with random

HYPOTHESIZED CURVE SHAPES IN GROWTH CURVE ANALYSIS

Table 3
Results of Each Model

Variables	Unconditional model			Model conditional on group status			Model conditional on group; fixed effects only for follow-up			Model conditional on group & intercept covariates		
	Estimate	<i>t</i>	<i>p</i>	Estimate	<i>t</i>	<i>p</i>	Estimate	<i>t</i>	<i>p</i>	Estimate	<i>t</i>	<i>p</i>
Intercept	6.79	11.2	<.001	7.95	10.1	<.001	7.95	10.1	<.001	7.86	10.2	<.001
Age										0.14	2.4	.018
Sex										0.78	0.9	.373
Co-morbidities										1.72	1.7	.095
Afternoon shift										2.38	1.9	.063
Night shift										-0.04	-0.0	.970
Delivery person										-0.77	-0.6	.554
Truck driver										-0.42	-0.2	.850
Sales representative										0.47	0.2	.864
Adjunctive care (<i>M</i> visits)										0.84	2.2	.035
Crossover group status				-1.89	-1.1	.296	-1.89	-1.1	.296	-4.00	-2.3	.027
Control group status				-2.63	-2.2	.030	-2.63	-2.2	.030	-3.59	-3.0	.004
Time slope	-0.56	-2.8	.006	-1.52	-7.7	<.001	-1.52	-7.7	<.001	-1.52	-7.7	<.001
Crossover*time slope				0.69	1.5	.525	0.69	1.5	.137	0.68	1.5	.140
Control*time slope				2.21	7.3	<.001	2.21	7.3	<.001	2.21	7.3	<.001
Follow-up (FU) slope	-1.28	-3.3	.001	-0.03	-0.1	.951	-0.01	-0.0	.979	-0.00	-0.0	.999
Crossover*FU slope				-3.03	-2.8	.005	-3.11	-2.9	.004	-3.15	-2.9	.004
Control*FU slope				-2.49	-3.6	.001	-2.46	-3.6	.001	-2.46	-3.6	.001
-2 * Log likelihood	1620.6			1557.7			1557.8			1519.6		
AIC	1634.6			1569.7			1565.8			1527.6		
BIC	1650.3			1583.2			1574.8			1536.6		
Variance of intercept	22.21	5.0	<.001	20.81	4.9	<.001	20.85	5.0	<.001	17.60	4.5	<.001
Variance of time slope	2.13	4.4	<.001	0.89	3.2	.001	0.90	3.4	.001	0.90	3.4	.001
Intercept*time slope	-3.99	-3.3	.001	-2.47	-2.8	.006	-2.51	-2.9	.004	-2.51	-2.9	.004
Variance of FU slope	2.74	1.5	.065	0.00	.							
Intercept*FU slope	1.01	0.5	.641	-0.57	-0.3	.766						
Time*FU slope	-1.74	-2.4	.015	0.05	0.1	.904						
Residual	3.85	7.8	<.001	3.77	9.3	<.001	3.77	9.3	<.001	3.76	9.3	<.001

Note. Intercept = Estimate of initial status; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; FU = follow-up.

intercepts and intervention-period slopes. The variance across participants in the follow-up period slope is not significant at $p < .05$. However, as these hypothesis tests may not be completely reliable (Singer, 1998) and since the p is small (.065), the random effects coefficient for the follow-up period is left in the model for now. The covariance term for intercept and time/slope (-3.99) indicates that individuals' baseline and intervention-period slope values are negatively related—i.e., higher (worse) baseline RDQ scores are associated with more negative slopes (greater improvement) during the intervention period. This is not unexpected as those with worse starting values tend to have more room to improve.

Conditional Model Results - Group Status

The results of the first conditional model are shown as the next set of columns in Table 3. Here the intercept and two slope values are estimated conditional on study group assignment (treatment or control group) and whether this participant volunteered for the crossover group. The intercept term has changed (from 6.79 to 7.95) because it now represents the average baseline RDQ for the treatment group. The average baseline value for the control group as a whole is lower at 5.32 (7.95 - 2.63). The treatment group improved an average of 1.52 RDQ points per month over the intervention period. Again, due to the asymptotic shape, two-thirds of this improvement happens by the second month. On average the control group's RDQ worsened by 0.69 (2.21 - 1.52) RDQ points per month. During the 3 months after the intervention period, the treatment group improved an additional 0.03 points on average from their intervention period end point, but this coefficient is not significant, indicating that for this group end-of-treatment outcome values held throughout the follow-up period.

The control group as a whole improved by an average of 2.52 (-2.49 + -0.03) RDQ points during the 3-month follow-up period. However, a number of those in the control group (the crossover group) received 1 month of treatment after the intervention period ended (during month 4). The crossover group improved by 3.03 RDQ points more than those in the control group who turned down crossover treatment. Because the crossover variable is centered on the control group mean, the average change during the follow-up period for the non-crossover group must be calculated as the average for the whole control group (2.52) minus the percentage of those in the control group who took the crossover treatment (42%) times the crossover group improvement (3.03) or an improvement of 1.25 RDQ points. The average change during the follow-up period for the crossover group is an improvement of 4.28 (1.25 + 3.03) RDQ points. This amount is the improvement due to the 1 month of treatment received and 2 months of post-crossover follow-up, and provides secondary

validation of the effectiveness of the study treatment. The crossover group also tends to have lower (better) baseline values of RDQ than the rest of the control group by 1.89 points, and this group had a greater worsening of RDQ over the intervention period (by an average of 0.69 RDQ points per month) than the rest of the control group. However, neither of these differences is statistically significant.

Again, the variance, covariance, and residual values (now *conditional* variance-covariance estimates) at the bottom of the table provide information on how much of the various sources of variation in individual RDQ values across participants has been explained by group variables. The conditional variance for the intercept (initial status RDQs) dropped by 6 percent from 22.21 to 20.81. Therefore, 6 percent of the variability across individuals in initial values is explained by group status. As treatment and control group status was randomly assigned, this percentage was not actually “explained” by group status. Instead it is the portion of variance across participants associated with group status due to a failure of randomization to equally allocate baseline RDQ scores. The conditional variance for the slope of the time variable during the intervention period dropped from 2.13 to 0.89. Therefore, group status explained more than half (58 %) of the variability across individuals’ intervention-period slopes. As the intention of the intervention was to change this slope, this is strong evidence of the effectiveness of this treatment. The conditional variance for the slope of the follow-up period time variable dropped to zero validating the non-significant estimate of variance from the unconditional model. As expected, the nested model comparison indicates that the fit of this model is significantly better than that of the unconditional model (χ^2 of 1620.6 - 1557.7 or 62.9 with 6 *df*, $p = <.001$).

Because this model validates the lack of random effects for the follow-up period slope (the variance of follow-up slope across participants is now zero), the model is re-run with only fixed effects for the follow-up period slope. These results are shown as the third model in Table 3. As can be seen the results stay essentially the same. The nested model comparison indicates that the fit of the model without random effects for the follow-up slope is not significantly different than that including random effects (χ^2 of 1557.8 - 1557.7 or 0.1 with 1 *df*, $p = 0.752$), and the AIC and BIC are reduced as an indication of the improved parsimony of this model.

Conditional Model Results - Group Status and Participant Characteristics

Akin to the process in general linear model fitting by which main effects are fitted before interactions, in HLM it is recommended to first develop the intercept model before that for the slopes (Raudenbush & Bryk, 2002). The fourth model shown in Table 3 includes all available

covariates for the intercept. The nine added explanatory variables have been centered on their grand means. Therefore, the group status coefficient estimates now represent those for a typical participant; one that represents the sample averages of age, sex, co-morbidities, adjunctive treatment use, and type and shift of worker. The average baseline RDQ score for the treatment group of 7.86 points differs from the 7.95 point estimate of the second model because it represents the baseline score for the treatment group after it has been adjusted to match the control group on the nine added variables. Similarly, note that the difference between the control and treatment group baseline scores is larger in this fourth model (3.59 vs. 2.63), as are the differences between baseline scores for the crossover group and the group that did not crossover (4.00 vs. 1.89)—both due to group matching.

The coefficients estimated for each of the nine new explanatory variables represent the change in the baseline RDQ score for a one-unit change in the variable. Baseline RDQ scores tend to be higher (worse) with age, for women, for those with co-morbidities at baseline, for those on the afternoon warehouse shift, for sales representatives, and for those using more adjunctive care at baseline. Baseline scores tend to be lower for night shift warehouse workers, delivery personnel, and truck drivers. The other coefficient estimates (those for the slopes of the intervention and follow-up periods) remain very similar to those of the third model, as expected.

Also as expected, the addition of covariates for the intercept has little effect on any of the conditional variance-covariance estimates other than the conditional variance for the intercept. It dropped to 17.60 indicating that 21 percent ($1 - 17.60/22.21$) of the variance in individuals' baseline RDQ scores is explained by group status and by available participant characteristics. Since 6 percent is explained by group status (above), 15 percent is explained by participant characteristics. No change is seen in the variance across participants in the slope of the intervention-period time variable. The nested model comparison indicates that the fit of this model, which allows participant characteristics to explain baseline scores, is significantly better (explains significantly more variance) than that of the model containing only group status (χ^2 of 1557.7 – 1519.6 or 38.1 with 9 *df*, $p < .001$).

Fully Saturated Conditional Model

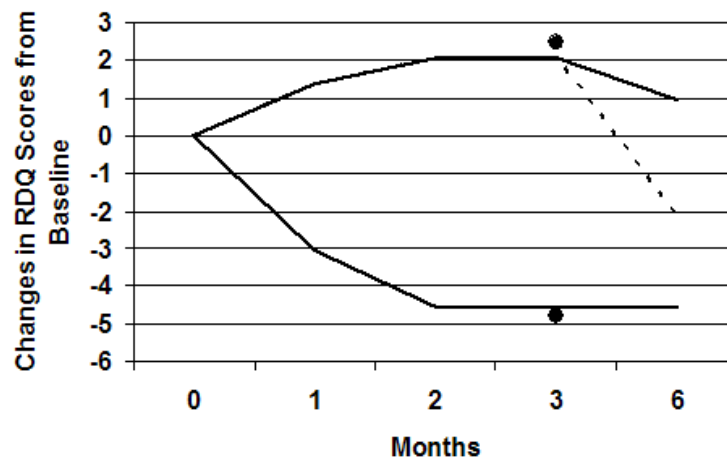
In order to test whether participants' characteristics also influenced their change in RDQ during the intervention and follow-up periods, a fully saturated model was tested. The nested model comparison for this model compared to that above indicates that the fully saturated model is not a significantly better fit to the data (χ^2 of 1519.6 – 1497.3 or 22.3 with 18 *df*, $p = .219$). Also, inclusion of all available covariates for both slopes did little

to help explain the variance across participants in intervention-period slopes—the percentage explained went from 58% to 61%. Therefore, the participant characteristics that could explain the remaining variation across participants in intervention-period slopes were not measured and are still to be discovered. The results of this model are not shown.

Comparison to Point Estimate Results

A straight comparison between the two group’s 3-month scores after adjusting for baseline values (as reported in Szczerko et al., 2007) yields a net improvement in the treatment group over the control group of 7.23 with a 95% confidence interval of -9.18 to -5.30. Using the results of the third model (Szczerko et al.’s analysis did not include covariates), a direct comparison between estimated 3-month RDQ scores (adjusting for baseline values) results in a net improvement in the treatment group over the control group of 6.63 (2.21 * 3). Figure 3 displays a graph of the growth curves estimated based on the third model and adjusted to be net of baseline. The graph also shows the point estimates from Szczerko et al. This graph illustrates not only the minor difference in the estimates attained, but also the wide difference in the amount of information generated by the two approaches. In one case two points are generated; in the other two curves and one line.

Figure 3. Change in Roland-Morris Disability Questionnaire (RDQ) scores from baseline for each group based on the growth curve analysis compared to the point estimates from Szczerko et al. (2007). The top line is the control group trajectory and splits into those that did not take the crossover treatment during the follow-up period (black line) and those that did (dashed line). The lower line is the treatment group trajectory.



The 95% confidence interval for the model-based estimate is -8.42 to -4.84 (based on a standard error for the slope of 0.91). Thus, utilizing the full dataset in the estimation of the 3-month impact of the intervention results in a narrowed confidence interval (a range of 3.58 points instead of 3.88 points). Because including baseline covariates did not change the intervention-period slope estimate, the results of the fourth model yield the same the net impact estimate (6.63 RDQ points) and confidence interval (95% CI: -8.42, -4.84).

Robustness of Model Estimates

The estimated coefficients for the fourth model were applied to the ODI and the pain VAS outcomes to test their robustness. No bias was found (the sum of the estimated residuals for each outcome is not significantly different from zero), and the model was found to have almost exactly the same fit to these data. The total variance explained by the model for all three variables is about 44 percent.

Discussion

This paper presents an example of growth curve analysis that incorporates an alternative approach to the modeling of non-linear growth curves: hypothesized, directly-coded curve shapes. This approach offers several benefits over orthogonal polynomial contrast coding, and “regular” polynomial coding (i.e., the inclusion of higher order time variables). First, the multicollinearity seen between time and the higher order time variables can be avoided with little cost to the interpretability of the coefficients. Second, in cases of a limited number of measurements per individual, non-linear shapes can be tested without losing degrees of freedom to higher-order time variables or contrast codes. And finally, this approach requires that specific alternative expected shapes be hypothesized prior to analysis allowing direct testing of the underlying theoretical basis of each.

Here the shape that best fit the data is one of rapid improvement followed by a slowing to a steady state (the asymptotic shape) that continued beyond the end of treatment and through the follow-up period, as hypothesized, for the treatment group. Validation of this theory (hypothesized shape) over the others as to how the intervention works across time indicates that this is a fast-acting treatment with effects that do not seem to relapse. This result might lead to consideration of shortening the duration of treatment and extending the follow-up period to look for possible later diminishing effects.

HYPOTHESIZED CURVE SHAPES IN GROWTH CURVE ANALYSIS

On the other hand, although the same intervention-period shape was the best fit for the control group, the slope coefficient was unexpectedly net positive (2.21 + -1.52) indicating an initial worsening slowing to a steady state. That the crossover group would improve after the intervention period when they received crossover treatment was expected. However, it was unexpected that the non-crossover members of the control group would also experience follow-up period improvement, albeit by a smaller amount (-2.46 + -0.00 - 0.42*-3.15 or 1.14 RDQ points). Because both alternative curve shapes and expectations for the slope coefficients (the negative sign of the intervention period coefficient and the non-significance of the follow-up-period coefficient) are specified ahead of time, the fact that something unexpected happened (the worsening of the control group during the intervention period and the subsequent improvement during the follow-up period for those not taking the crossover treatment) is highlighted. In the absence of prior specification, this result may have been lost in the focus on the net impact of the treatment. Instead these results call for a re-evaluation of the control intervention.

In addition to the benefits of hypothesized directly-coded growth curve shapes, the use of growth curve analysis itself provides a better estimate of the impact of a treatment intervention than point estimates, as evidenced by the narrower confidence interval. This increase in accuracy is accompanied by an increase in information regarding not just end-point impacts, but of the healing trajectories of participants over the entire study period. As discussed above, this information is crucial to future practice recommendations and research designs.

The ability to partition the unexplained variance into baseline, slope, and residual components is also useful to the process of model estimation and interpretation. Using this information it can be shown that available participant characteristics (age, sex, co-morbidities, adjunctive care, and work type and shift) explained 15% of the variance across participants in initial status. Group status explained 58% of the variance across participants in rate of change. Although the majority of variation in initial status remains to be explained by future work, it is noteworthy that the largest part of the variation across participants in RDQ change over time—the variation that is the target of treatment—is explained by whether they received the study treatment or not.

In summary, the use of growth curve analysis in the estimation of the impact of study interventions not only improves the statistical quality of the estimates, but also provides information on the healing trajectory—information that is vital to the determination of appropriate treatment duration. The use of hypothesized growth curve shapes improves standard growth curve methodology by requiring a priori specification and allowing explicit testing of viable shape options while maximizing model degrees of

freedom and avoiding multicollinearity. As funding for intervention studies is limited, it is crucial that the best estimates are obtained and the most information gained from each study conducted. The use of hypothesized directly-coded curve shapes and growth curve analysis can help ensure that none of these data are wasted.

Author Note: Patricia M. Herman, RAND Health, RAND Corporation; Lee Sechrest, Department of Psychology, University of Arizona. This publication was made possible by Grant Number F32 AT003363 from the National Center for Complementary and Alternative Medicine (NCCAM). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NCCAM, or the National Institutes of Health. The authors would like to acknowledge Drs. Orest Szczurko, Kieran Cooley, and Edward Mills of the Canadian College of Naturopathic Medicine for the use of the data from their clinical trial. Correspondence concerning this article should be addressed to Patricia M. Herman, RAND Corporation, 1776 Main Street, Santa Monica, CA 90407-2138. email: pherman@rand.org

References

- Beurskens, A. J. H. M., de Vet, H. C. W., & Koke, A. J. A. (1996). Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain, 65*, 71-76.
- Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9*, 30-52.
- Fairbank, J. C., Couper, J., Davies, J. B., & O'Brien, J. P. (1980). The Oswestry low back pain disability questionnaire. *Physiotherapy, 66*, 271-273.
- Feldman, H. A. (1988). Families of lines: random effects in linear regression analysis. *Journal of Applied Physiology, 64*, 1721-1732.
- Heo, M., Faith, M. S., Mott, J. W., Gorman, B. S., Redden, D. T., & Allison, D. B. (2003). Hierarchical linear models for the development of growth curves: an example with body mass index in overweight/obese adults. *Statistics in Medicine, 22*, 1911-1942.
- Huskisson, E. C. (1982). Measurement of pain. *Journal of Rheumatology, 9*, 768-769.
- Kihlstrom, J. (2005, September). 'Patients' and 'clients'. *Observer, 18*, 6.
- Kurdek, L. A. (1999). The nature and predictors of the trajectory of change in marital quality for husbands and wives over the first 10 years of marriage. *Developmental Psychology, 35*, 1283-1296.
- Lipsey, M. W. (1990, April). *Theory as method: small theories of treatments*. Paper presented at the Research Methodology: Strengthening Causal Interpretations of Non-experimental Data, Washington.

HYPOTHESIZED CURVE SHAPES IN GROWTH CURVE ANALYSIS

- Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, 5, 23-43.
- O'Connell, A. A., & McCoach, D. B. (2004). Applications of hierarchical linear models for evaluations of health interventions: demystifying the methods and interpretations of multilevel models. *Evaluation & the Health Professions*, 27, 119-151.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111-163.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Roland, M., & Morris, R. (1983). A study of the natural history of low-back pain, part I: development of a reliable and sensitive measure of disability in low back pain. *Spine*, 8, 141-144.
- Szczurko, O., Cooley, K., Busse, J. W. et al. (2007). Naturopathic care for chronic low back pain: a randomized trial. *PLoS ONE*. 29, e919.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth curve models. *Journal of Educational and Behavioral Statistics*, 24, 323-355.
- West, S. G., Wu, W., & Ryu, E. (2006). Growth curve models: an introduction, *Western Psychological Association 86th Annual Convention*. Palm Springs, CA: Western Psychological Association.