Modeling Longitudinal Count Data : Testing for Group Differences in Growth Trajectories Using Average Marginal Effects

Sarah Mustillo, Lawrence R. Landerman and Kenneth C. Land

Sociological Methods & Research 2012 41: 467 originally published online 29 June 2012
DOI: 10.1177/0049124112452397

Article

Sociological Methods & Research
41(3) 467–487

DOI: 10.1177/0049124112452397
http://smr.sagepub.com

Modeling Longitudinal Count Data: Testing for Group Differences in Growth Trajectories Using Average Marginal Effects

Sarah Mustillo[1], Lawrence R. Landerman[2], and Kenneth C. Land[3]

Abstract

To test for group differences in growth trajectories in mixed (fixed and random effects) models, researchers frequently interpret the coefficient of Group-by-Time product terms. While this practice is straightforward in linear mixed models, it is less so in generalized linear mixed models. Using both an empirical example and synthetic data, we show that the coefficient of Group-by-Time product terms in a specific class of mixed models—mixed Poisson models for count outcome variables—estimates the group difference in slope as the multiplicative change with respect to the baseline rates, not differences in the predicted rate of change between groups. The latter can be obtained from computing the marginal effect for the expected

[1]Department of Sociology and Center for Aging and Life Course, Purdue University, West Lafayette, Indiana
[2]Center for the Study of Aging and Human Development , School of Nursing, Duke University
[3]Department of Sociology and Center for Population Health and Aging, Population Research Institute, Duke University

Corresponding Author:
Sarah Mustillo, Purdue University, Department of Sociology, 700 W. State St. West Lafayette, IN 47907, USA.
Email: smustillo@purdue.edu.

response with respect to time by group following model estimation. We propose and illustrate the use of marginal effects to test and interpret group differences in rate of change over time following estimation with mixed Poisson regression models.

Keywords

Group differences, growth curve models, marginal effects, count data

Introduction

Longitudinal panel research designs in which an initial wave of respondents or subjects is surveyed, interviewed, or observed repeatedly at subsequent intervals of time are increasingly used in the social sciences. Across various disciplines, the linear mixed (fixed and random effects) model (LMM) has become the standard procedure for analyzing these types of data when the research objective is to model trajectories of change over time. A large body of literature exists to guide researchers in the theory, specification, estimation, and postestimation evaluation techniques of LMMs; however, many outcomes of interest to social scientists are categorical. While standard, general use, statistical software programs have begun to provide routines for non-LMMs or generalized LMMs (GLMMs), less attention has been paid by researchers to how to properly specify, estimate, test, and interpret these models. Typically, researchers use the same techniques that are available for LMMs, often misinterpreting the results.

With an LMM, testing differences in trajectories between or among groups is straightforward. Researchers typically include a cross level product term between group membership (such as gender) and time to examine group differences in slope, or change over time, of the dependent variable. For example, Singer and Willett (2003:73) state that the z statistic from such an interaction term in an LMM serves as a significance test for group differences in rate of change. Others have suggested that this same procedure can be used in non-LMMs. For example, Rabe-Hesketh and Skrondal (2005) note that the coefficient of the product term can be interpreted as indicating group differences in the rate of change over time for the log odds in logistic models (pp. 115-18) and ordinal models (pp.155-61). In this case, the z statistics or t statistic is then a test of the log-transformed outcome, but it may be more useful to consider a test on the untransformed scale.

Several prior publications have noted that interpreting interaction terms in *cross-sectional nonlinear models* is not straightforward. In a linear model, the interaction effect is the cross partial derivative of the expected value of the response variable with respect to the two variables involved in the interaction, say $X1$ and $X2$. In this case, the coefficient of the interaction term is the same as the marginal effect; hence the coefficient itself can be interpreted as the group difference in the effect of time on the dependent variable with the corresponding Z test as the significance test. In nonlinear models, the interaction effect is also the cross partial derivative of $E(Y)$ with respect to $X1$ and $X2$, but, in this case, it is not the same as the coefficient of the interaction term (Ai and Norton 2003; Corneliβen and Sonderhof 2009; Norton, Wang, and Ai 2004). Further, Rothman (2002) and Rothman, Greenland, and Lash (2008) have suggested that when the dependent variable is dichotomous, estimates of whether the effect of one risk factor varies across levels of another can produce misleading results and that interactive effects should be assessed on an *additive* scale (effects on disease prevalence) rather than on a multiplicative scale (effects on the odds of disease).

This problem of specification and interpretation of interaction effects extends to the *generalized linear model in a longitudinal context* wherein the interaction term of a grouping variable and a time variable is a test for the group difference in slope on a ratio scale with respect to baseline values. The difference in the expected rate of change, rather than the ratio of change, is measured not by the coefficient of the interaction term but by taking the derivative of the conditional expectation with respect to time by group (as demonstrated below for a specific nonlinear panel model). When the baseline differences in rates are small, both quantities provide similar results. Conversely, when the baseline differences are larger, the coefficient for the interaction effect and the marginal effect will provide different results and answer different questions. In clinical trials, where randomization eliminates baseline differences and clinical importance is based on a ratio-scaled change, a ratio scale may be preferable (Keene 1995). This is not the case with survey data where a risk factor hypothesized to affect post-baseline change will typically have exerted a similar effect prior to baseline and produced baseline differences.

In this article, we demonstrate the nature and scope of this issue for count data with a specific instance of a GLMM, a mixed Poisson model, using both an empirical example and synthetic data. We demonstrate the use of marginal effects to test for group differences in slopes and make recommendations for future analyses.

Table1. Unadjusted Mean Number of IADLs by Cognitive Impairment Status at Baseline for Four Panel Waves/Time Points, EPESE Data

	Panel wave			
	1	2	3	4
	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)
No impairment	0.16 (0.63)	0.21 (0.77)	0.24 (0.84)	0.39 (1.03)
Impairment	0.94 (1.59)	1.10 (1.81)	1.23 (1.81)	1.45 (1.86)

Cognitive Impairment and Change in Disability: A Panel Study

To illustrate the nature of the issue, we first present an empirical example using data from the Duke site of the Established Populations for Epidemiological Studies of the Elderly (EPESE) study to examine difference in changes of a count measure of disability by baseline cognitive status among 4,066 respondents aged 65 and older across four waves of data collection (Cornoni-Huntley et al. 1990). The Duke EPESE sample from this multisite collaborative epidemiological investigation of older adults is based on five contiguous counties in central North Carolina. Respondents were asked whether they could perform each of five standard instrumental activities of daily living (IADLs) tasks (traveling, shopping, preparing meals, doing housework, and managing finances) without assistance (Fillenbaum 1985). The dependent variable is coded as a count of activities the respondent cannot do without assistance, that is, a higher number means more disability. Table 1 presents group means on IADLs disability over four years by cognitive status (cognitively impaired vs. cognitively intact) at the initial baseline survey.

Based on prior incidence studies (e.g., Dodge et al. 2005; Moritz, Kasl, and Berkman 1995; Raji et al. 2002), we expect yearly increases in IADLs disability will be greater in the cognitively impaired group. Consistent with this expectation, the data show an overall increase of .23 IADLs in the cognitively intact, or .06 per year, and .51 in the cognitively impaired, or .13 per year (Table 1). Therefore, it appears that the cognitively impaired increase at approximately twice the yearly rate of the cognitively intact over time. Because we suspect significant respondent-specific changes from wave to wave in IADLs disability by cognitive impairment that can be treated as random departures from overall group-specific changes, we turn to a GLMM to perform the test.

Mixed Poisson Regression Model

We analyze the longitudinal data on number of IADLs at each wave for each respondent with a GLMM using the Poisson family and a log link. The specific form of the GLMM applied is variously referred to as a *random coefficients model* or a *growth curve model*. It includes subject-specific intercepts and slopes, which allow individuals to vary on their initial level of cognitive impairment and their rate of change over time. In a growth curve model, time-specific individual-level measures are assumed to contain input from two sources: the process under consideration and random error or departure from group-specific mean changes that can be modeled with individual-specific intercept and slope coefficients across time.

Preliminary data analysis of the outcome variable Y_{it}, the count of IADLs disabilities for panel respondent i at time t, is distributed as a rare event that can be modeled as a Poisson variable with conditional expectation for individual i as a function of the time period of observation $\text{Time}_{it} = 0, \ldots, 3$ given by (Dobson and Barnett 2008:166):

$$E(Y_{it}) = \mu_{it} = \beta_{0i} \exp(\beta_{1i}\text{Time}_{it})$$

where we specify that both Poisson distribution function parameters, β_{0i} and β_{1t}, may vary across individual respondents. Following conventional generalized linear model specifications and using the logarithmic link function $\ln(\mu_{\text{it}})$, the conditional expectation function specifying the mixed Poisson regression model at level 1 then is:

$$\ln(\mu_{it}) = \beta_{0i} + \beta_{1i}\text{Time}_{it}$$

The level 2 submodel is:

$$\beta_{0i} = \beta_0 + \beta_2\text{Impairment}_i + \zeta_{0i}$$
$$\beta_{1i} = \beta_1 + \beta_3\text{Impairment}_i + \zeta_{1i}$$

and the composite model is:

$$\ln(\mu_{it}) = \beta_0 + \beta_1\text{Time}_{it} + \beta_2\text{Impairment}_i + \beta_3\text{Impairment}_i * \text{Time}_{it} + \zeta_{0i} + \zeta_{1i} * \text{Time}_{it}$$

where ln denotes the natural logarithm, μ_{it} is the mean response variable (expected number of IADLs) for individual i at time t; β_0 is an overall intercept, Time_{it} is a level-one variable for wave or year of the survey with fixed-effect regression parameter β_2, Impairment_i is a level-two respondent-specific variable denoting respondent i's impairment status at the baseline survey with fixed-effect coefficient β_2, $\text{Impairment}_i \times \text{Time}_{it}$ is the cross level product or interaction term for the change in the effect of respondent i's baseline impairment status on the outcome variable at panel wave t with fixed-effect coefficient β_3, and ζ_{0i} and ζ_{1i} are residuals around the population-averaged intercept and slope, respectively. The distribution of the random effects is assumed to be bivariate normal with variances ψ_{11} and ψ_{22} and covariance ψ_{21}. Since the log likelihood for this model has no closed form, this model can be estimated using direct numerical integration, various quasi-likelihood approaches, or Bayesian analysis. In this analysis, we estimate this model with xtmepoisson in Stata 12 (Statacorp 2011) which uses direct numerical integration by adaptive Gaussian quadrature.

This conditional expectation function models expected changes in log counts for respondent i at wave t. Estimates of coefficients can be interpreted as representing the difference between the log of expected counts from values X_k to $X_k + 1$ of regressor X. Because the difference in two logs is equal to their quotient, it is also the log of the ratio of expected counts. Exponentiating gives the ratio of the expected counts or the *incidence rate ratio* (IRR), wherein a one unit increase in X is associated with an increase or decrease in the ratio of expected counts by a factor of $\exp(b_k)$.

As shown in Table 2, the IRR for cognitive impairment, or the estimated rate ratio at baseline, indicates that the cognitively impaired have a rate or IADL count 16.73 times higher than the cognitively intact at baseline. The IRR for time or wave of the panel indicates an increasing trajectory over time, such that with all other variables held constant, we can expect a yearly increase in the rate ratio of 1.72 IADLs for the cognitively intact. The IRR for the interaction term (IRR = 0.83, SE = 0.04) indicates that the cognitively impaired experience a *smaller* increase over time compared to the cognitively intact.

The expected or predicted counts of IADLs from the estimated mixed Poisson regression model show a pattern opposite to the IRR for the interaction term but similar to the unadjusted means, with yearly *increases in the predicted number of IADLs greater in the impaired group* (0.31 vs. 1.03 overall and 0.078 per year vs. 0.258 per year for the cognitively intact and cognitively impaired, respectively). To summarize, the unadjusted means and the model-based predicted counts show that the impaired experience a

Table 2. Estimated Mixed Poisson Model of Number of IADLs Regressed on Cognitive Impairment by Time, EPESE Data

Fixed parameters			
	B	*SE*	IRR
Cognitive impairment	2.817***	(0.161)	16.73
Time	0.541***	(0.060)	1.72
Cog impairment × Time	−0.188***	(0.046)	0.83
Intercept	−4.555***	(0.144)	
Random components			
Slope variance	0.102	(0.0187)	
Intercept variance	7.562	(0.679)	
Covariance	−0.487	(0.115)	
Summary statistics			
N	15,016		
Chi-square	434.050		
Log likelihood	−8,346.699		

Note: Standard errors in parentheses. IADLs = instrumental activities of daily living; IRR = incidence rate ratio.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

greater increase in problems with IADLs over time. The model IRR indicates the opposite—that the intact subjects increase at a greater rate than the impaired subjects because the IRR of the interaction terms measures the multiplicative effects with respect to the baseline rate for each group (cf., Buis 2010). If researchers are more interested in overall group differences in the rate of change rather than change as a percentage of the baseline rate, the interaction term does not address this interest.

Marginal Effects

To test the group difference in rate of change over time rather than ratio of change, we seek a significance test to determine whether we can reject the null hypothesis of no group differences in the trajectory of additive change for the population. A variety of authors have developed methods for cross-sectional and case–control data (Assman et al. 1996; Hosmer and Lemeshow 1992; Knol et al. 2007; Li and Chambless 2007; Norton et al. 2004). Here, we provide a method for repeated-measures count outcomes by using a significance test of group-based marginal effects following a mixed Poisson model with a Group-by-Time interaction term.

The marginal effect is an approximation of the amount of change of expected Y per unit change in a predictor and is based on taking the derivative or partial derivative of $E(Y)$ with respect to an explanatory variable, X. Marginal effects can help to summarize model results, particularly with generalized linear models. In a linear model, the derivative of $E(Y)$ with respect to X is the same as the coefficient of X. That is, the marginal effect is the slope or the amount of change in $E(Y)$ per one unit change in X and is the same as the model coefficient and the same regardless of the values of any other explanatory variable in the model provided there are no multiplicative terms (see Greene 2003 for a general treatment). For a generalized linear model like the Poisson model, because of the nonlinearity in the relationship between X and Y, the marginal effect is not the same as the regression model coefficient but is a function of the coefficient that depends in part on the type of model being estimated (Kennedy 2003). In this case, the marginal effect varies based on the value of X itself and the values of the other explanatory variables in the model (Long 1997).

Given that the marginal effect varies based on the values of the other explanatory variables, these values must be set in some way to compute the effect. Often, this is done by setting the other explanatory variables to their means called *marginal effects at the mean* or MEMs. Alternatively, a marginal effect can be computed for each case, using the subjects' observed values on the explanatory variables. The average of these marginal effects is then computed and is thus referred to as *average marginal effects* or AMEs. Both MEMs and AMEs have been widely used outside of sociology, but most recently scholars seem to favor AMEs (Cameron and Trivedi 2009). Because MEMs are computed at mean values of explanatory variables, they can be based on values that do not exist in the data set or in reality. For example, when setting gender to the mean to compute MEMs, the value will be neither 0 nor 1, but a fraction equal to the mean of male or female in the sample (say .55). Since no one can be 55% male or female, the resulting marginal effect is based on a nonsensical value. By computing a marginal effect for every subject in the data set using their observed values on explanatory variables and then averaging across subjects, AMEs avoid this problem (see Wooldridge 2009 for a more detailed explanation of marginal effects in general and MEMs vs. AMEs in particular).

Computing the marginal effect for $E(Y)$ with respect to X in a generalized linear model can help explain the model results, even in a model with only main effects, but in a model with interaction or multiplicative terms, marginal effects are even more useful. As Ai and Norton (2003) explain, in a linear model, the interaction effect is the cross derivative of $E(Y)$ with respect to the interacted variables, say $X1$ and $X2$, if both variables are continuous, or the discrete difference if both variables are dichotomous. In both cases, this is equal to the coefficient of the interaction term. In a generalized linear model, the interaction effect is still the cross partial derivative, but this is not the same as the coefficient for the interaction term in this case. Fortunately, we can compute the cross partial derivative with existing software via Stata's margins postestimation command. Because the Group-by-Time interaction term is a Dichotomous-by-Continuous product term, we will compute the derivative or marginal effect of $E(Y)$ with respect to time separately by group first and then test for a significant difference between the two values with a Wald test second (StataCorp 2009).

Synthetic Example

Having identified and presented the issue with a motivating example from the EPESE data above, we now turn to a synthetic example to further explicate the difference between the ratio of change and the rate of change and show how and why the interaction term provides a test of one but not the other. We use generated data to illustrate the difference and circumstances under which it occurs. Further, we demonstrate that testing differences in the marginal effect or derivative of the conditional expectation is an appropriate method to test for group differences in the overall expected rate of change as opposed to the ratio of change. We estimate the marginal effects and perform such a test using standard software. Putting aside IADLs for now, we created three random Poisson-distributed response variables with a fixed difference in slope over time but a different mean in the EPESE data set. Specifically, we generated one Poisson-distributed outcome variable with a mean of 4, one with a mean of 5, and one with a mean of 6, but all with the same gradually increasing slope for males and a more steeply increasing slope for females (see Online Appendix A for graphs which can be found at http://smr.sagepub.com/supplemental/).

In each case, the total increase in counts on the synthetic variables was approximately 2 in males and 4 in females, but the intercept (e.g., baseline value) varied by the mean of the distribution. The objective was to hold the gender difference in change over time constant, but successively increase the mean by 1, to demonstrate how the coefficient and IRR for the Group-by-Time product term changed as the baseline values changed, even though the additive gender difference in change over time remained the same.

Following the mixed Poisson model with the Group-by-Time interaction term, AMEs were calculated using margins in Stata, which is based on the following equations:

$$\frac{\partial E(Y)}{\partial T} = \hat{P} = \frac{1}{w.} \sum\nolimits_{i=1}^{n} (\delta_i(S_p)) w_i h(Z_i, \tilde{\boldsymbol{\theta}})(*)$$

with

$$w. = \sum\nolimits_{i=1}^{n} (\delta_i(S_p)) w_i$$

and marginal effects for each panel respondent based on her or his observed values on the explanatory variables are:

$$h(Z_i, \hat{\theta}) = \frac{\partial e^{z_i \tilde{\boldsymbol{\theta}}}}{\partial T}$$

where $\tilde{\boldsymbol{\theta}}$ is a vector of estimated regression coefficients from the mixed Poisson model, Z_i is a vector of covariate values for individual *i, T* is time, $\delta_i(S_p)$ is an indicator variable denoting whether individual *i* is in subpopulation S_p, w_i is the weight for the *i*th observation if performing a weighted analysis with user-specified weights, and *n* is the sample size (StataCorp 2009). If the analysis is unweighted, $w_i = 1$. In brief, expression (*) indicates that the AME for the entire panel, $\hat{P}$, is calculated by first calculating the AME $h(Z_i, \tilde{\boldsymbol{\theta}})$ for each panel respondent, which then is weighted by w_i, sorted into the subpopulation S_p to which the respondent belongs by the $\delta_j(S_p)$ indicator variable, summed across the *n* panel members, and averaged by dividing by the sum of the sorted panel weights, *w*. The variance of $\hat{P}$ is estimated using the Delta method.

After estimating the full mixed model equation for each generated dependent variable, we then used this expected values equation to calculate the AMEs of time (panel wave) for each category of the grouping variable, including all subjects' observed values on covariates in the calculation and then taking the sample average. Hence, in this case, the marginal effects are the average predicted rate when the grouping variable is 0 (males) and when it is 1 (females). We expected the product term to be the most inconsistent with the predicted counts and marginal effects when the mean of the Poisson variable was lowest and to be more consistent with the data as the mean increased, because change in the ratio scale is more similar to change in the margins at higher values given the same gender difference.

Table 3 presents the results from the three synthetic count response variables described above. Because the means differ, the intercepts and end points for the trajectories differ by model despite the fact that the gender difference in the slopes is the same. In model 1, with mean of 4 for the synthetic response variable, the IRR for the Group-by-Time product term is less than 1 and significant, indicating a smaller change over time for the female group (IRR = 0.937, $p < .01$), whereas the difference in marginal effects is positive and significant (diff = 0.667, $p < .01$), indicating a larger change over time for females. The difference in findings between the IRR and the marginal effects is due to the ratio of change from wave to wave being larger for men than for women, despite the fact that the overall rate of change is twice as high in the females compared to the males.

Specifically, in model 1, the average predicted count for males on the synthetic dependent variable goes from 1.10 at baseline to 1.55 at the first follow-up to 2.18 at the next follow-up to 3.06 at the last wave, for a mean ratio of change equal to approximately 1.4 (e.g., 1.55/1.10 = 1.41; 2.18/1.55 = 1.41; 3.06/2.18 = 1.40). The corresponding values for females are 3.11, 4.09, 5.40, and 7.12 for a mean ratio of change equal to approximately 1.32 (e.g., 4.09/3.11 =1.32; 5.40/4.09 = 1.32; 7.12/5.40 = 1.32). Dividing the female ratio, 1.32, by the male ratio, 1.40 (the dreaded ratio of ratios!), equals .94, which is roughly equivalent to the IRR of the product term in the mixed Poisson model (given rounding error).

On an additive scale, the difference in average predicted count for males is 0.04 between baseline and first follow-up, 0.63 between first follow-up and second follow-up, and 0.89 between second follow-up and last follow-up, for a total difference between baseline and third follow-up of 1.96. The corresponding values for females are 0.99, 1.3, and 1.71, with a total difference of 4.01. Subtracting the male total value from the female total value equals 2.05, which is roughly equivalent to the difference in marginal effects multiplied by time ($0.667 \times 3 = 2.001$), given rounding error.

In model 2, with an increase in the mean to 5, the IRR in the model is approximately 1 and nonsignificant, indicating no gender differences in change over time (IRR = 1.006), but the difference in marginal effects is positive, significant, and almost identical to the difference in marginal effects from model 1 (diff = .665, $p < .01$). In this case, because the ratio of change from wave to wave is the same for men and women, the IRR shows no effect, but the marginal effects reflect the overall gender difference in slope. Lastly, in model 3, as the mean of the variable increases again by one unit to 6, the IRR for the product term is positive and significant, because the ratio of change is lower (IRR = 1.03, $p < .01$). The difference in

Table 3. Mixed Poisson Regression Models Estimated for Generated Outcome Variables in EPESE Data (n = 16,648)

	Model 1			Model 2			Model 3		
Mean Outcome =	4			5			6		
	B	(*SE*)	IRR	*B*	(*SE*)	IRR	*B*	(*SE*)	IRR
Time	0.340***	(0.008)	1.406***	0.222***	(0.006)	1.250***	0.165***	(0.059)	1.180***
Female	1.037***	(0.020)	2.822***	0.671***	(0.016)	1.957***	0.498***	(0.013)	1.646***
Female × Time	−0.065***	(0.009)	0.937***	0.005	(0.007)	1.006	0.029***	(0.006)	1.030***
Intercept	0.080***	(0.019)		0.741***	(0.014)		1.397***	(0.011)	
Chi-square	13,824.89			11,435.03			9,775.27		
Log likelihood	28,298.13			31,527.75			34,031.66		

Note: Standard errors in parentheses. IRR = incidence rate ratio.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

marginal effects for this model is consistent with the previous two models (diff = 0.642, $p < .01$) and, in this case, consistent with the IRR for the interaction term.

Examining the predicted counts in Table 4 helps to explain the results. In model 1, additive increases in the predicted count are smaller among males ($3.06 - 1.10 = 1.96$) and greater among females ($7.12 - 3.11 = 4.01$). On a ratio scale, these additive differences translate into a larger increase among males ($3.06/1.1 = 2.78$) and a smaller increase among females ($7.12/3.11 = 2.29$). For this model, the IRR is less than 1, reflecting that the rate ratio for males is greater than the rate ratio for females ($1.32/1.40 = 0.93$). This is the case despite the fact that overall change is greater among females. In model 2, overall change remains the same among males (an increase of about 2) and among females (an increase of about 4); however, because the overall mean on Y is increased by 1 and the baseline values increase, change on a ratio scale is now 1.24 among males and 1.25 among females, and the IRR is about 1, reflecting this equality. In model 3, gender-specific overall increases remain the same while the mean is again increased by 1. Due to this larger mean and even higher baseline values, the wave to wave ratio of change is larger for females than for males, resulting in an IRR greater than 1. The IRR for this model is the only one that is consistent with our expectations and consistent with the way the gender interaction was constructed.

In sum, while overall differences in the predicted counts are essentially the same across models, only in the third model the rate ratio of female to male becomes positive. Conversely, the marginal effects are similar in all three models, positive and significant. When multiplied by time, the marginal effects in each model equal approximately 2 for males and 4 for females, as expected. With a mixed Poisson model, the product term is assessing group differences in rate ratios, and these differences may or may not be consistent with the overall group differences in the slope of change from wave to wave. Both methods are assessing change, but each can give very different answers, depending on the nature of the distribution and the group difference itself. If researchers are interested in assessing group differences in the overall rate of change, not the ratio of change with respect to the group-specific baseline values, the IRR for the product term is not the appropriate statistic. Computing the marginal effect, or, the derivative of the conditional response with respect to time for each level of the grouping variables, and computing a significance test on the differences will better answer research questions about group differences in rates of change over time.

Table 4. Predicted Counts From Mixed Poisson Regression Models Estimated for Generated Outcome Variables With a Gender Interaction

	Model 1						Model 2						Model 3					
	Male			Female			Male			Female			Male			Female		
Time	P^a	Diff	Ratio	P^a	Diff	Ratio	P^a	Diff	Ratio	P^a	Diff	Ratio	P^a	Diff	Ratio	P^a	Diff	Ratio
0	1.10			3.11			2.08			4.06			3.08			5.06		
1	1.55	0.04	1.40	4.09	0.99	1.32	2.60	0.52	1.24	5.10	1.04	1.25	3.63	0.55	1.17	6.15	1.09	1.22
2	2.18	0.63	1.40	5.40	1.30	1.32	3.25	0.65	1.24	6.42	1.31	1.25	4.28	0.65	1.17	7.48	1.33	1.22
3	3.06	0.89	1.40	7.12	1.71	1.32	4.06	0.81	1.24	8.06	1.65	1.25	5.05	0.77	1.17	9.09	1.61	1.22
3 versus 0		1.96	2.78		4.01	2.29		1.97	1.95		4.00	2.00		1.97	1.64		4.03	1.80
Fem ratio/ Male ratio		0.94						1.01						1.03				
dy/dt																		
Male		0.693*** (0.018)						0.659***(0.021)						0.687***(0.027)				
Female		1.359***(0.021)						1.325*** (0.022)						1.329*** (0.025)				
Difference		0.667***(0.028)						0.665***(0.031)						0.642***(0.037)				

Note: [a]Predicted count.

Taken together, the results in Tables 1 to 4 show that interpreting a multiplicative product term as to determine whether the within-subject rate of change varies by a between-subject factor should be done with caution and only in the case when the researcher is interested in multiplicative rather than additive change. On a substantive level, in the first set of analyses, the multiplicative pattern in our sample suggested that cognitive impairment was protective for subsequent increases in IADL disability. This conclusion was inconsistent with previous findings (Dodge et al. 2005; Moritz et al.; Raji et al. 2002) and with the observed changes in the rate as shown in Table 1. Several researchers have proposed alternative methods for testing group differences in cross-sectional logistic regression models, either by using an additive scale (Rothman 1998) or by using the differences in predicted probabilities at each time point (Long 2009). In the case of a longitudinal model, however, testing for group differences in the predicted probabilities at each time point, while informative, is not the same as a significance test for overall slope differences. Therefore, given the above issues associated with using the product term to compare subgroup differences on a multiplicative scale, we return to our motivating example to test for differences in the rate of change of IADLs by cognitive impairment using marginal effects rather than the IRR for the Group-by-Time product term.

Table 5 presents again the results from the mixed Poisson model analysis of the EPESE data. Results show a significant and negative Group-by-Time product term (IRR = 0.830, $p < .01$), but a significant and positive group difference in marginal effects (diff = 0.093, $p < .01$), indicating that the average increases in the rate of disability are significantly greater among the cognitively impaired. The results from the significance test on the difference in marginal effects reflect the differences in the predicted counts based on the mixed Poisson model analysis. For the cognitively intact group, the marginal change in Y with respect to time shows a modest increase in the predicted counts of disability between baseline and time 3, while for the cognitively impaired group, increases across time are greater in magnitude.

In model 2, to expand the analysis and demonstrate the use of marginal effects with additional explanatory variables in the model, we add age, ethnicity, gender, marital status, and income as covariates (code available in online Appendix B which can be found at http://smr.sagepub.com/supplemental/). We also add a product term between Marital Status and Time, in addition to the product term between Cognitive Impairment and Time, as preliminary evidence showed a possible slope difference in IADLs by marital status such that unmarried subjects appeared to increase in disability more steeply over time compared to married subjects. Of the two product

Table 5. Estimated Mixed Poisson Models of IADLs Regressed on Cognitive Impairment Status at Baseline With Marginal Effects for Interaction Terms, EPESE data

	Model 1			Model 2		
	B	*SE*	IRR	*B*	*SE*	IRR
Time	0.541***	(0.102)	1.716***	0.564***	(0.082)	1.758***
Cognitive impairment	2.817***	(2.686)	16.727***	2.130***	(1.299)	8.413***
Cog impairment × Time	−0.187***	(0.038)	0.830***	−0.186***	(0.035)	0.830***
Age				0.114***	(0.008)	1.121***
Female				−0.032	(0.113)	0.969
Black				0.225*	(0.131)	1.253*
Income				−0.029***	(0.006)	0.972***
Married				0.005	(0.152)	1.005
Married × Time				−0.025	(0.040)	0.976
Intercept	−4.555***			−12.701***		
dy/d*t*						
No cog impairment	0.015***	(0.002)		0.022***	(0.003)	
Cog impairment	0.108***	(0.018)		0.163***	(0.024)	
Difference	0.093***	(0.017)		0.141***	(0.023)	
Married				0.019***	(0.003)	
Unmarried				0.052***	(0.006)	
Difference				0.033***	(0.005)	
N	15,016			15,016		
Chi-square	435.434			783.838		
Log likelihood	−8,344.206			−8,174.471		

Note: Random coefficients omitted. IADLs = instrumental activities of daily living; IRR = incidence rate ratio.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

terms, only Cognitive Impairment by Time is statistically significant in the regression model (IRR = -0.186, $p < .01$), but we still calculate the difference test on the marginal effects, as a nonsignificant product term does not necessarily mean that there is no significant difference in marginal effects. In this case, the difference tests of marginal effects for both cognitive impairment and marital status are significant (cognitive impairment diff = 0.141, $p < .01$; marital status diff = 0.033, $p < .01$). Had we only interpreted the coefficient of the interaction terms, we would have concluded that cognitive impairment is protective and that marital status has no impact on change over time in disability level.

Discussion and Conclusion

Using EPESE data on baseline cognitive impairment and changes in IADL disability over time, we showed that when a repeatedly measured outcome is a count variable assessing group differences in change over time is not straightforward. Analyses using a cross level product term from a Poisson model indicated that baseline cognitive impairment was protective for subsequent increases in IADL disability, despite the fact that increases in the predicted count of disabilities showed the opposite pattern. Especially counterintuitive was a substantial and significant negative product term when the rate of disability in the cognitively impaired group increased by 65% over 3 follow-ups compared with a 41% increase among the cognitively intact. The product term assessed whether the ratio of change varied with respect to baseline values, which, in this case, did not address the question of interest. To assess group differences in the overall slope differences, a different approach was needed.

We illustrated the nature of the problem using three synthetic-dependent variables with identical gender differences in slope but different means and hence different baseline values and different ratios of change. We found that, despite identical gender differences in slopes for each of the three outcome variables, only one product term reflected that overall difference. The rate ratio in the model with the lowest mean was negative, indicating a significant and negative relationship. In the second model, the overall difference in change over time was the same as the first model, but the IRR was approximately 1 because the ratio of change between males and females was the same. In the third example, the mean of the dependent variable was higher and the rate ratio was positive, thus agreeing with the overall gender difference.

To address the need for a significance test of additive change, we presented and illustrated an alternative method of testing for slope differences which used a generalized linear mixed Poisson model with an interaction term in an initial step and then the derivative or partial derivative in the expected response (e.g., marginal effects) with respect to time by group as a second step. Finally, we tested for group differences in that rate of change using a standard Wald test. Results based on the alternative method were consistent across models in both the synthetic examples, the motivating IADL example, and the expanded IADL model. In the expanded IADL model, results confirmed that cognitive impairment was associated with an increase in disability over time and also detected a significant increase in slope of disability by marital status, even when the product term for marital status and time was not significant in the mixed Poisson model. Hence, the marginal effect method addressed our research question while the cross level product term did not.

To our knowledge, ours is the first study to demonstrate the complexity of testing change over time in a mixed Poisson model. Given this, we selected a relatively simple model for change to demonstrate as clearly as possible how and why this problem can arise and to illustrate an alternative analytic method. As a result, we have not addressed how this same problem may manifest itself and can best be addressed in more complex GLMM models, including those with continuous between-person factors, nonlinear trajectories of change, and different distributions (e.g., negative binomial) on the dependent variable; nor is it clear whether discrepancies between ratio-based and marginal measures of change are as problematic in randomized designs where initial group differences on the outcome measure are not present. We believe that these limitations can and should be addressed in future studies.

A further limitation is that the marginal effects presented here have the potential to oversimplify complex relationships. A nonlinear (logit, probit, Poisson) model implies a distribution of interactions on an additive scale, where a predictor's marginal effects can be different at each level of each predictor in the model (Norton et al. 2004). In our analyses, diverse within-group (impairment) marginal effects are averaged to estimate between-group differences in marginal effects. An examination of the within-group marginal effects, which might further explain why the between-group differences are present, would require a more complex set of analyses than those proposed here.

When researchers are interested in group differences in change over time with count-dependent variables, they must decide whether they are interested in assessing change on a ratio scale or evaluating change in terms of predicted rate of change over time. We have presented evidence to show that a product term from a Poisson model should not be used to determine whether the within-person rate of change varies across levels of a between-person predictor when one is interested in the rate of change over time rather than the ratio of change. As an alternative, we demonstrated the use of a difference test of marginal effects from a mixed Poisson model. In our data, this alternative method provided plausible and meaningful results when a Poisson product term did not. Results based on this approach were more congruent with the observed data and better addressed the research question. Simulations based on a variety of different rates and interactive patterns are needed to determine the utility of this approach in more situations. Nevertheless, the results here suggest the use of marginal effects following model estimation with an interaction term, rather than the coefficient of the interaction term itself, may be preferable for assessing whether the rate of within-person change varies across between-person factors at this time.

Declaration of Conflicting Interests

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Lawrence R. Landerman acknowledges support from the National Institutes of Health, National Institute on Aging, Claude D Pepper Older Americans Independence Center; Grant No. P30 AG028716.

References

Ai, Chunrong and Edward C. Norton. 2003. ‘‘Interaction Terms in Logit and Probit Models.’’ *Economics Letters* 80:123-29.

Assmann SF, Hosmer DW, Lemeshow S, Mundt KA. 1996. Confidence intervals for measures of interaction. Epidemiology 7: 286-90.

Buis, Maarten L. 2010. ‘‘Stata Tip 87: Interpretation of Interactions in Non-linear Models.’’ *The Stata Journal* 10:305-8.

Cameron, A. Colin and Pravin K. Trivedi. 2009. *Microeconomics Using Stata*. College Station, TX: Stata Press.

Corneliβen, Thomas and Katja Sonderhof. 2009. ‘‘Partial Effects in Probit and Logit Models with a Triple Dummy Variable Interaction Term.’’ *The Stata Journal* 9:571-83.

Cornoni-Huntley, Joan C., Dan G. Blazer, Dwight B. Brock, and Mary E. Farmer, eds. 1990. *Established Populations for Epidemiological Studies of the Elderly: Volume II. Resource Data Book*. NIH Publication No. 90-495. Washington, DC: Public Health Service. National Institutes of Health..

Dobson, Annette J. and Adrian G. Barnett. 2008. *An Introduction to Generalized Linear Models*. 3rd ed. New York, NY: Chapman & Hall.

Dodge, Hiroko H., Kadowaki Takashi, Hayakawa Takehito, Yamakawa Masanobu, Sekikawa Akira, and Ueshima Hirotugu. 2005. ''Cognitive Impairment as a Strong Predictor of Incident Disability in Specific ADL-IADL Tasks among Community-Dwelling Elders: The Azuchi Study.'' *Gerontologist* 45:222.

Fillenbaum, George C. 1985. ''Screening the Elderly: A Brief Instrumental Activities of Daily Living Measure.'' *Journal of the American Geriatric Society* 33:698-706.

Greene, William H. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.

Hosmer DW, Lemeshow S. 1992. Confidence Interval Estimation of Interaction. Epidemiology. Vol. 3, no. 5: 452-456.

Keene, Oliver. 1995. ''The Log Transformation is Special.'' *Statistics in Medicine* 14:811-9.

Kennedy, Peter. 2003. *A Guide to Econometrics*. 5th ed. Cambridge, MA: MIT Press.

Knol MJ van der Tweel I, Grobbee DE Numans M, Geerlings M. 2007. Estimating interaction on an additive scale between continuous determinants in a logistic regression model.

Li, R. and Chambless, L. 2007. Test for additive interaction in proportional hazards models. Ann. Epidemiol., 17 (3), 227-236.

Long, J.S., 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

Long, J. Scott. 2009. ''Group Comparisons in Logit and Probit using Predicted Probabilities.'' Unpublished manuscript. Retrieved (http://www.indiana.edu/~jslsoc/files_research/groupdif/groupwithprobabilities/groups-with-prob-2009-06-25.pdf).

Moritz, Deborah J., Stanislav V. Kasl, and Lisa F. Berkman. 1995. ''Cognitive Functioning and the Incidence of Limitations in Activities of Daily Living in an Elderly Community Sample''. *American Journal of Epidemiology* 141:41-9.

Norton, Edward, Hua Wang, and Chunrong Ai. 2004. ''Computing Interaction Effects in Logit and Probit Models.'' *The Stata Journal* 4:103-16.

Rabe-Hesketh, Sophia and Anders Skrondal. 2005. *Multilevel and Longitudinal Modeling Using Stata*. College Station, TX: Stata Corp.

Raji Mukaila, A., Glenn V. Ostir, Kyriakos S. Markides, and James S. Goodwin. 2002. ''The Interaction of Cognitive and Emotional Status on Subsequent Physical Functioning in Older Mexican Americans: Findings from the Hispanic

Established Population for the Epidemiologic Study for the Elderly.'' *Journal of Gerontology: Medical Sciences* 57:M678-82.

Rothman, Kenneth J. 1998. *Modern Epidemiology*. Philadelphia, PA: Lippincott-Raven.

Rothman, Kenneth J. 2002. *Epidemiology: An Introduction*. New York, NY: Oxford University Press.

Rothman, Kenneth J., Sander Greenland, and Timothy W. Lash, eds. 2008. *Modern epidemiology*. 3rd ed. Philadelphia, PA: Lippincott-Willians and Wilkins.

Singer, Judith D. and John B. Willet. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York, NY: Oxford.

Statacorp. 2011. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.

Wooldridge, Jeffrey. 2008. *Introductory Econometrics: A Modern Approach*. 4th ed. Mason, OH: South-Western.

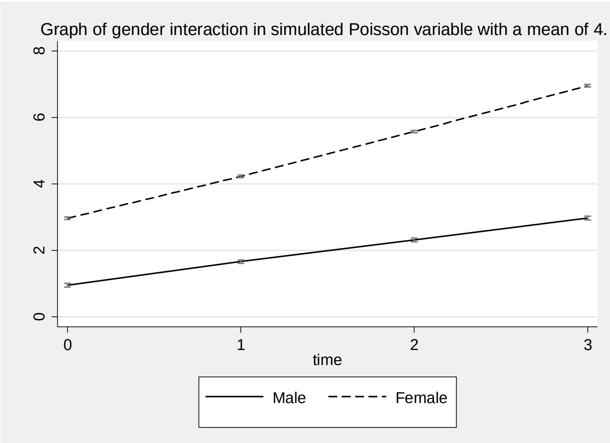
Bioss

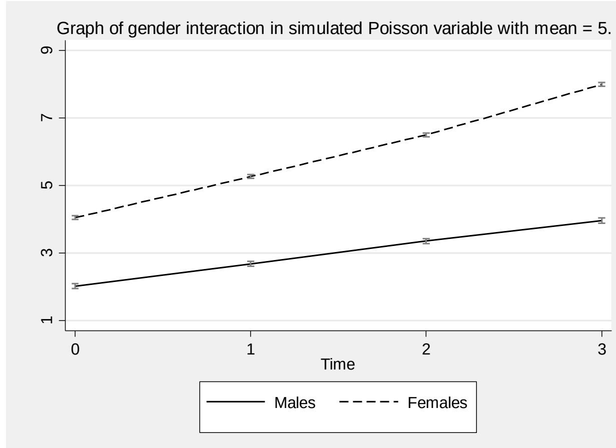
Sarah Mustillo is an associate professor of sociology and faculty fellow of the Center for Aging and Life Course at Purdue University. Her methodological research interests include various aspects of missing data and the longitudinal modeling of categorical data, while her substantive interests focus on the effects of different types of childhood adversity on child and adolescent physical and mental health trajectories.

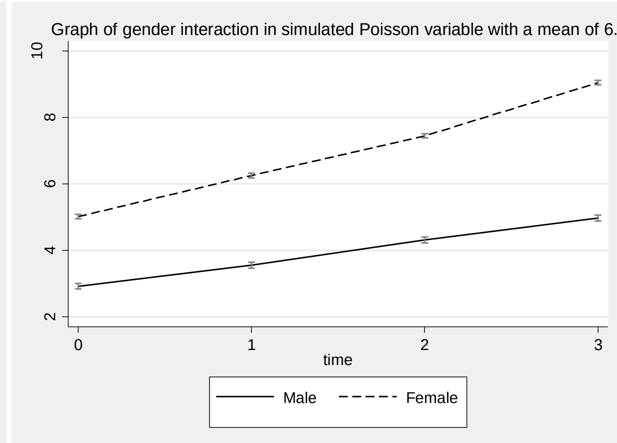
Lawrence R. Landerman is an Associate Professor of Medical Research at the Center for the Study of Aging and Human Development at the Duke University Medical Center. His research interests include social stratification and health outcomes, the intersection of biological and social factors related to successful aging, and statistical models for change over time.

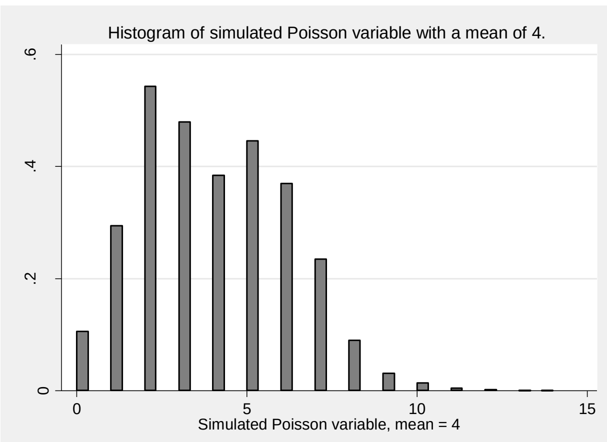
Kenneth C. Land is the John Franklin Crowell Professor of Sociology and Demographic Studies at Duke University. He is an elected Fellow of the American Statistical Association and was the 1997 recipient of the Paul F. Lazarsfeld Award from the Methodology Section of the American Sociological Association. His research interests are in the development of mathematical and statistical models and methods for substantive applications in demography, criminology, and social indicators/quality-of-life studies.

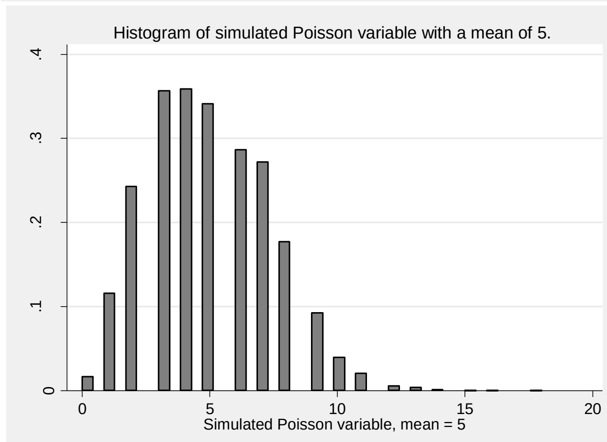
Appendix A. Graphs of synthetic variables and gender differences in slope.

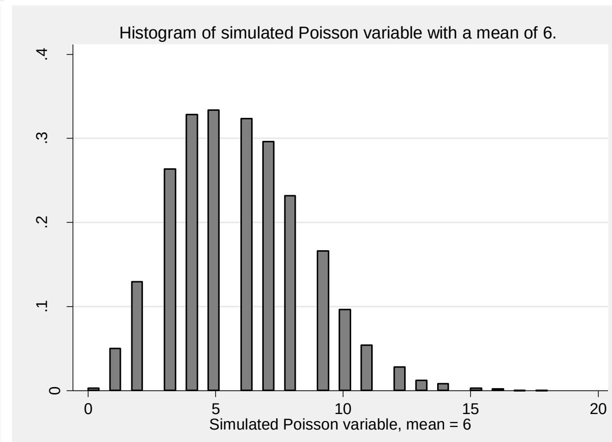












Appendix B: Stata code for Table 5, Model 2.

```
*Model 2

xtmepoisson badl c.time##i.cogimp c.age i.female i.black i.married##c.time midinc ||person:time,

mle var cov(unstr)

margins , over(married) dydx(time) post predict(fixedonly)

lincom _b[0.married] - _b[1.married]


xtmepoisson badl c.time##i.cogimp c.age i.female i.black i.married##c.time midinc ||person:time,

mle var cov(unstr)

margins , over(cogimp) dydx(time) post predict(fixedonly)

lincom _b[0.cogimp] - _b[1.cogimp]
```