Published in final edited form as:

Soc Sci Res. 2011 September 1; 40(5): 1456–1464. doi:10.1016/j.ssresearch.2011.05.006.

MODELING REPEATED MEASURES OF DICHOTOMOUS DATA:

Testing Whether the Within-Person Trajectory of Change Varies Across Levels of Between-Person Factors

Lawrence R. Landerman^{*}, Sarah A. Mustillo^{**}, and Kenneth C. Land^{***}

^{*}Center for the Study of Aging and Human Development, School of Nursing, Duke University

^{**}Department of Sociology, Purdue University, West Lafayette, Indiana

^{***}Department of Sociology and Center for Population Health and Aging, Population Research Institute, Duke University

Abstract

In this paper, we consider the following question for the analysis of data obtained in longitudinal panel designs: How should repeated-measures data be modeled and interpreted when the outcome or dependent variable is dichotomous and the objective is to determine whether the within-person rate of change over time varies across levels of one or more between-person factors? Standard approaches address this issue by means of generalized estimating equations or generalized linear mixed models with logistic links. Using an empirical example and simulated data, we show (1) that cross-level product terms from these models can produce misleading results with respect to whether the within-person rate of change varies across levels of a dichotomous between-person factor; and (2) that subgroup differences in the rate of change should be assessed on an additive scale (using group differences in the effects of predictors on the *probability* of disease) rather than on a multiplicative scale (using group differences in the effects of predictors on the *odds* of disease). Because usual approaches do not provide a significance test for whether the rate of *additive* change varies across levels of a between-person factor, sample differences in the rate of additive change may be due to sampling error. We illustrate how standard software can be used to estimate and test whether additive changes vary across levels of a between-person factor.

Research designs in which an initial wave of respondents or subjects is observed repeatedly over time are increasingly used in the social sciences. The primary goal of these studies is to depict change over time and to identify factors that influence the direction and rate of change. These factors can include time constant (e.g. gender) or time-changing (e.g., marital status) variables. As Molenberghs and Verbeke (2006:7) have observed, mixed (hierarchical, multilevel) models have become the main tool for the analysis these kinds of data. Examples of studies across various disciplines using these methods include the influence of language exposure on early vocabulary growth (Huttenlocher et al. 1991), the impact of early teacher effectiveness on trajectories of student achievement (Palardy and Rumberger 2008), how age, sex, race, class, and place of residence affect patterns of delinquency, substance abuse and health problems among young persons (Elliot, Huizinga, and Menard 1989), the

© 2011 Elsevier Inc. All rights reserved.

Address all correspondence to Lawrence R. Landerman, Center for the Study of Aging and Human Development, Box 3003, Duke University Medical Center, Durham, NC 27710-3003; lrl@geri.duke.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

relationship between education on change in blood pressure over the life course (Loucks et al. 2011), and the effects of age and race on trajectories of self esteem among older persons (Shaw, Liang, and Krause 2010).

Bodies of codified knowledge and methodological guidelines for the statistical analysis of repeated measures in such panel designs have developed correspondingly in various disciplines and fields of research (e.g., Molenberghs and Verbeke 2006; Hsiao 2003; Singer and Willett 2003; Walls and Schafer 2006; Wooldridge 2002). In this paper, we consider how these repeated-measures data should be modeled and interpreted when the dependent variable is dichotomous and the objective is to determine whether the within-person rate of change over time varies across levels of a between-person factor (or group). With a linear model and a continuous dependent variable, a cross-level product term can be used to estimate and test differences in the rate of change between groups (Singer and Willett 2003).¹ Some have suggested that a cross-level product term from a logistic model can be used in a similar manner to examine group differences in the trajectory of change when the dependent variable is dichotomous (e.g., Molenberghs and Verbeke 2006:282–287; Rabe-Hesketh and Skrondal 2005:115–118).

Using both an empirical example and simulated data, we show that using a cross-level product term from a logistic model to evaluate group differences in the rate of change can produce highly misleading results, especially when substantial group differences in baseline prevalence are present. We argue that subgroup differences in the rate of change over time should be assessed on an additive scale (using group differences in the effects of predictors on the *probability* of an outcome) rather than on a multiplicative scale (using group differences in the effects of predictors on the *odds* of an outcome). Because standard approaches do not provide an overall estimate or significance test for whether the additive change varies across subgroups in the population, we illustrate how marginal effects on the probability can be estimated based on a logistic model, and then used to estimate and test whether additive changes in the probability vary with baseline status.

Baseline Cognitive Impairment and Change in IADL Disability

To illustrate the nature of the problem, we focus on the analysis of a particular outcome variable in a specific longitudinal panel study. We study changes in a dichotomous measure of disability among elderly respondents followed over three yearly waves after an initial baseline survey. Data are from the Duke University site of the multisite National Institute on Aging (NIA)-funded Established Populations for Epidemiologic Studies of the Elderly (EPSE) program. This was a 6-year annual (1986/87–1992/93) longitudinal study of community residents aged 65 years or older. A four-stage sampling design was used to obtain a probability sample of 4162 persons (80% of those contacted), aged 65 years or older, living in households in five contiguous counties in central North Carolina. Our analyses are based on 4066 non-proxy respondents. (Non-proxy respondents gave their own responses, whereas proxy respondents were those whose responses were provided by another person – typically a relative – due to a health-related inability to respond.) For a more complete description of the sampling design, see Blazer et al. (1991).

The substantive issue of interest is whether the rate of increase in IADL disability over time is greater among those who were cognitively impaired at baseline compared to those who were cognitively intact. Respondents were asked whether they could perform each of five

¹A 'cross level' product term in a multilevel model is constructed by multiplying a level 1 within-person predictor (here, time) by a level 2 between-person predictor (cognitive impairment). In linear models, the product term corresponds to the interaction effect for the two predictors involved. In nonlinear (logit, probit) models it does not (Norton et al. 2004). For this reason, we do not refer to this term as an 'interaction' term or 'interaction' effect.

standard instrumental IADL tasks (traveling, shopping, preparing meals, doing housework, managing finances) without assistance (Fillenbaum 1985). In the analyses presented below, respondents unable to perform four or five tasks were classified as disabled. Similar results were obtained in analyses using different cut-points on the disability measure. Based on Pfeiffer (1975), cognitive impairment is defined as making eight or more errors on the short portable mental status questionnaire (SPMSQ) with one additional error allowed for those with a grade school education or less, and one fewer errors allowed for those with greater than high school education. Based on prior incidence studies (e.g., Moritz, Stanislav, and Berkman 1995; Raji et al. 2002; Dodge et al. 2005), we expect yearly increases in IADL disability will be greater in the cognitively impaired group.

In Table 1, changes in the percent disabled are consistent with this expectation, with yearly increases over the four waves of the study averaging 8.3% among the cognitively impaired compared with 6.3% among the cognitively intact. As we will show below using generalized estimating equations (GEE's) and generalized linear mixed models (GLMM's), results based on a product term from a logistic model can lead to a very different conclusion. The data in Table 1 also show 60% attrition over three follow-ups. To the extent that data are not missing completely at random (MCAR) our GEE results will be biased, and our GLMM-based coefficients will be biased to the degree that the data are not missing at random (MAR) contingent upon the independent variables in our models (Schafer and Graham 2000). Because our purpose is methodological rather than substantive, this source of bias should not adversely affect our interpretations and conclusions.

To simplify our discussion of logistic product terms, we restrict our sample in Table 2 to the first two waves of the EPESE data. On an additive scale, time-related increases in the probability of disability are greater among those who were cognitively impaired at baseline (.11 vs. .04). On a multiplicative scale, time-related changes in the odds of disability show an opposite pattern with increases in disability greater among those who are cognitively intact at baseline (2.07 vs. 1.59). The multiplicative product term (.77) in the bottom row of Table 2 is based on a logistic model where disability is regressed on cognitive impairment, time and cognitive impairment by time. The product term is negative (<1) because changes in the odds of disability are smaller in the cognitively impaired group. The additive product term (.07) is derived from a linear probability model with the same predictors and product term. The product term is positive because changes in the probability of disability are greater in the impaired group. (We are aware of the statistical limitations of the linear probability model (Hanushek and Jackson 1977). It is used here to illustrate the difference between an interaction assessed on an additive scale and an interaction assessed on a multiplicative scale.)

In Table 2, larger changes in the probability of disability among the impaired translate to smaller changes in the odds because changes in the odds are assessed on a multiplicative scale relative to baseline, and baseline disability is substantially higher in the impaired group (.36 vs. .04). As a result, the multiplicative product term is negative. The hypothetical data in Table 3 further illustrate the impact of baseline prevalence on an odds-based product term. For the low risk group, baseline prevalence is .05 and the probability of disability increases by 5% at follow-up. For the first high risk group, baseline prevalence remains at .05 while the probability of disability increases by 10%. Converting to a ratio scale, the odds of disability increase by 3.32 per year in the first high risk group compared to 2.09 in the low risk group, resulting in a positive multiplicative product term of 1.59. With high risk group 2, we keep yearly change in the probabilities the same while increasing baseline prevalence to .10. This increase in the prevalence reduces the odds-based product term to 1.08. When baseline prevalence is increased to .30 in high risk group 3, the odds-based product term becomes negative (.74). *In general, given the same percent increases in*

disability over time, increases in baseline prevalence in the high risk group drive the odds-based product term in a negative direction. The additive product term, on the other hand, remains constant as baseline prevalence increases in the high risk group.

We next use all four waves of EPESE data to estimate two longitudinal models which are commonly employed with repeated measures data and a dichotomous outcome. These are generalized estimating equations (GEE) with a logistic link, and a generalized linear mixed model (GLMM) with a random intercept and a logistic link. For each, dichotomous IADL disability is regressed on baseline cognitive impairment, time, and a cognitive impairment by time product term. The GEE model was estimated with SAS PROC GENMOD; the GLMM with SAS PROC GLIMMIX (SAS 2005).

Generalized estimating equations incorporate dependence among repeated observations via a user-specified working correlation matrix which allows for correlations on the dependent variable over time (Liang and Zeger 1986; Twisk 2004). Coefficients estimated with GEE will tend to be close to those estimated with logistic regression, but standard errors will be corrected for dependency. We use an unstructured working correlation structure to estimate the marginal probability of an event as:

$$\ln(p_{ij}/1-p_{ij})=\beta_0+\beta_1 Time_{ij}+\beta_2 IP_i+\beta_3 IP_i * Time_{ij}, \quad (1)$$

where:

p_{ij} denotes the probability that survey respondent i is IADL disabled at time j ,

$Time_{ij}$ is a time of survey wave counter variable for respondent i and wave $j=0, 1, 2$, or 3,

IP_i is respondent i 's impairment status at the baseline survey, and

the β_k are regression coefficients to be estimated.

The GEE - based coefficients in Table 4 are very close to the logistic results in Table 2. Because baseline disability is higher in the impaired group, the product term again indicates that increases in disability are significantly greater in the intact group (OR=.82, $p<.01$), while increases in the predicted probabilities are greater in the impaired group (.12 vs. .08). The GEE model provides a significance test for the odds based multiplicative product term but not for whether additive changes in the probabilities differ by baseline impairment.

A GLMM model takes the form:

$$\ln(p_{ij}/1-p_{ij})=\beta_0+\beta_i+\beta_1 Time_{ij}+\beta_2 IP_i+\beta_3 IP_i * Time_{ij} \quad (2)^2$$

where β_i represents a random, subject-specific intercept which is allowed to vary across respondents, and other coefficients are as in model 1.

²Equation 2 presents the GLMM model in composite form which can be decomposed into the Level-1 (within person) and Level-2 (between person) submodels as follows:

$$\ln(p_{ij}/1-p_{ij})=\beta_{0i}+\beta_{1i} Time_{ij} \quad (3.1)$$

$$\beta_{0i}=\gamma_{00}+\gamma_{01} IP_i+u_{0i} \quad (3.2.1)$$

$$\beta_{1i}=\gamma_{10}+\gamma_{11} IP_i+u_{1i} \quad (3.2.2)$$

In Table 4, the GLMM coefficients for time are larger than the GEE, and about equal across levels of impairment. As a result, the odds ratio for the impairment by time product term is now close to one (1.03) and non-significant. On an additive scale, increases in the probability of disability are here much greater among the impaired (.27) than among the intact (.03), but this difference is not tested by the GLMM model.

The GLMM coefficients differ from their GEE counterparts because they are statistically corrected for attenuation due to “unobserved heterogeneity” (Allison 1999a:76–78). Logistic and GEE coefficients are attenuated toward zero to the degree that all relevant causes are not included as predictors. Because measures of all relevant causes are almost never available, unobserved heterogeneity (between-person residual variation) can be substantial and attenuation, which increases with heterogeneity, can be substantial as well. The random effect in Table 4 (15.0, $p < .01$) is an estimate of unobserved heterogeneity, and represents the combined effects of all omitted covariates that are related to disability. By explicitly incorporating it into the regression model, we adjust for unobserved heterogeneity and correct for attenuation.

With the inclusion of a random intercept, $(\Delta Y/\Delta t)$ is estimated separately for each individual and then averaged, and the GLMM coefficients reflect change at the individual level. As a result, Coefficients based on GEE and GLMM have different interpretations. GLMM provides ‘*subject-specific*’ coefficients which measure change at the individual level, while GEE models provide ‘*population averaged*’ coefficients which indicate the average change in the overall prevalence associated with a one-unit change in a predictor for the population as a whole. Each GLMM coefficient is the product of a logistic coefficient and a correction factor which increases with the amount of unobserved heterogeneity³. Holding the correction factor constant, the GLMM impairment by time interaction term will decrease as baseline disability in the high risk group increases and the logistic coefficient decreases. Hence, while the multiplicative product term in our GLMM is no longer negative in this example, it is still affected by group differences in baseline disability.

Summary of Problems Identified

The preceding analyses have shown that cross-level product terms in logistic models can be strongly influenced by group differences in baseline prevalence rates. Bollen (1989, ch 4) and others warn against using standardized regression coefficients ($b \cdot S_x/S_y$) to compare the effects of regression predictors across subgroups because group differences in metric regression effects are affected by group differences in S_x and S_y . With the use of multiplicative product terms, group differences on baseline prevalence similarly affect regression coefficients, and create similar problems. When the baseline rate of disability is larger in the group that is hypothesized to be at greater risk, multiplicative product terms are driven in a negative direction. In longitudinal studies, risk factors for subsequent adverse events will typically exert similar effects prior to baseline. As a result, the pattern observed here – higher baseline disability in the high risk group – will be the rule rather than the exception in studies of health and disability, multiplicative product terms will be biased in a negative direction, and can indicate a protective effect when changes in the probabilities suggest an opposite pattern (see also Hoetker 2004).

Group differences in additive change over time are not similarly affected by baseline disability rates (Table 2). Given this, and the problems identified for a multiplicative product term, we believe that whether the rate of within-person change over time is greater in one

³When the random component is normally distributed, the GLMM coefficient is approximated as a function of the GEE coefficient and a variance term where: $B_{GLMM} = B_{GEE} * \{ [.346 \text{var}(B_j) + 1] \}$ and $\text{var}(B_j)$ is the unobserved heterogeneity (Allison 2005:66).

group than another (or across levels of a between-person factor generally), should be measured on an additive scale. Our position is consistent with a consensus among epidemiologic researchers. Based on incongruous findings like those presented above, and on Rothman's highly influential "sufficient component" cause model (1998), they conclude that whether the effect of one risk factor on a dichotomous outcome variable varies across levels of another should be assessed on an additive scale (Kaufman 2007; Rothman, Greenland, and Lash 2008; Ahlbohm and Alfredsson 2005). More recently, VanderWeele and Robins (2008) have derived empirical tests for the presence of sufficient component interactions involving binary measures, and VanderWeele (2009) shows how these relate to linear, logistic and log-linear models.

Moving from a multiplicative to an additive test for whether the effect of one risk factor varies across levels of another will also avoid a problem with the logistic product term discussed by Norton and colleagues (2004). They show that, while the ($X1 \times X2$) product term in a linear model is the interaction effect -- the difference in the effect of $X1$ on Y across levels of $X2$ -- this is not the case with nonlinear (logit, probit) models. For these models, the interaction effect is the cross-partial derivative of the expected value of Y with respect to $X1$ and $X2$ (Norton et al. 2004, Cornelissen and Sonderhof 2009).

Testing Whether Additive Change in the Probability of Disability Varies Across Subgroups

The predicted probabilities from the GEE and GLMM models in Table 4 indicated that increases in the probability of disability are greater among the cognitively impaired *in our sample data*. We require 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 testing whether the effect of one risk factor on a dichotomous outcome varies across levels of another on an additive scale for cross-sectional, cohort, and case-control data (Hosmer and Lemeshow 1992; Assman et al. 1996; Knol et al. 2007; Li and Chambless 2007; Norton et al. 2004). We extend their work by providing methods which can be used with longitudinal repeated-measures data.

While a longitudinal linear probability model could be used to test whether additive change in the probability of disability varies across subgroups, the standard errors would be incorrect due to the presence of heteroscedasticity. Further, fitting a straight line to the probability of an event miss-specifies the shape of the underlying probability distribution, and can generate predicted probabilities less than zero or greater than one (Hanushek and Jackson 1977). A more statistically sound alternative is to use the results from a longitudinal logistic model to test the difference in the marginal effect of one predictor across levels of another. The marginal effect is an approximation of the amount of change in expected Y per unit change in a predictor *on an additive scale* -- the derivative of the mean expected probability of Y given X . It can be estimated with existing software via the Stata 'margins' postestimation command (StataCorp 2009). Several recent technical documents illustrate how the margins command can be used to estimate and test differences in the marginal effect of one predictor across levels of another (Buis 2010; Stata FAQ 2010a; Stata FAQ 2010b; Ender 2010).

We will illustrate the use of the marginal effect to test whether the effect of time varies with baseline cognitive impairment on an additive scale for both GEE and GLMM models with logistic links⁴. The GLMM approach is generally preferred because it corrects for heterogeneity shrinkage and because it provides unbiased parameter estimates when the data are missing at random (Molenberghs and Verbeke 2006; Allison 2005). However, both models are often used in practice, and concern that GLMM corrections rest on unverifiable

assumptions which (if incorrect) can seriously bias the results has led some to conclude that the GEE model is the safer alternative (Hubbard et al. 2010).

To make our example more realistic, we included age, gender, race, and education as covariates. For GEE, we used an unstructured residual matrix. For GLMM we used Gaussian quadrature points which are thought to perform better than alternative pseudo-likelihood-based techniques in correcting for attenuation (Rodrigues and Goldman 1995; Pinheiro and Chao 2006). Because quadrature -based iterations can stop at a local minimum, we re-estimated the model increasing the number of quadrature points, and stopped at 20 when fixed and random effects remained stable with further increases. Because differential heterogeneity can confound comparisons of logistic coefficients across subgroups (Allison 1999b), we also tested for differences in intercept variance (heterogeneity) by cognitive impairment. This test was positive ($p < .01$), and we modified the GLMM to allow for separate residual variance estimates for the intact and the impaired.

With the method we propose, marginal effects (dY/dt) are calculated by taking the derivative of $E(Y)$ with respect to time separately by group and then testing for a significant difference using a Wald test. Within each level of the grouping variable, marginal effects are averaged across subjects whose values (dY_i/dt_i) reflect their values on the covariates included in the model. For each subgroup, the marginal effect represents the average change in $E(Y)$ per unit change in time, where $E(Y)$ is the expected probability of disability. To make use of State's margins command we first re-estimated our GEE and GLMM models using Stata's `xtgee` and `xtmelogit` commands.⁵ (Logistic results estimated earlier with SAS PROC GENMOD and SAS PROC GLIMMIX were identical to those obtained using `xtgee` and `xtmelogit`).

In Table 5, group differences in the marginal effects (dy/dt) test whether increases in disability vary with cognitive impairment on an additive scale, and reflect changes in the probability rather than the odds of subsequent disability. With GEE, the predicted probability of disability increases by .158 among the cognitively impaired, compared with .072 among the intact, resulting in a difference of .086. With GLMM, the corresponding difference is .214. Both differences are positive and significant, indicating that the cognitively impaired are at increased risk for subsequent disability. The larger difference for the GLMM suggests that group differences in the individual-level trajectories are substantially larger than differences in the population-averaged effects. The fact that the intercept variance was substantially larger among the impaired led to a larger adjustment for attenuation in this group, and a larger difference in the marginal effects. The marginal effects are consistent with, but not the same as average changes in the predicted probabilities

⁴Because the logistic model is most commonly used in relevant applications, we used a logistic link at step 1. As the logit and probit distributions differ only slightly and only in the tails of the distributions, results based on a probit model at step 1 should be nearly identical to those reported.

⁵STATA code for GEE and GLMM models (Table 5) is as follows
 where: diadl=IADL disability; time=time; zcogimp=baseline cognitive impairment.
 /* GEE MODEL (dropping ns zcogimp*time interaction)
 xtgee diadl aage black female educ i.zcogimp c.time, i(person) fam(bin) link(logit) corr(unstr)
 estimates store model1
 estimates restore model1
 . margins,over(zcogimp) dydx(time) post
 . lincom _b[1.zcogimp] - _b[0.zcogimp]
 /*GLMM MODEL*/
 xtmelogit diadl aage black female educ i.zcogimp##c.time, ///
 ||person: zcogimp, noconstant ||person: notcog, noconstant intpoints(20) refineopts(iterate(0))
 estimates store model1
 estimates restore model1
 margins,over(zcogimp) dydx(time) post predict(fixedonly)
 lincom _b[1.zcogimp] - _b[0.zcogimp]

at the bottom of Table 5. Note that with the GEE model, we were able to estimate and test for an interactive effect on an additive scale despite removing the non-significant multiplicative product term from our logistic model. This is the case because a (logistic) model which is additive in the logged odds can be interactive in the probabilities.

Discussion and Conclusions

Using EPESE data on baseline cognitive impairment and changes in IADL disability over time, we showed that when a repeatedly-measured outcome is dichotomous, using a product term from a logistic model to assess whether the rate of within-person change varies across subgroups can give substantively implausible and misleading results (i.e., that baseline impairment is protective for subsequent disability). This is the case because a logistic product term measures changes in the odds of an event on a ratio scale, and because ratio-based measures of subgroup change are strongly influenced by group differences in baseline prevalence rates. Given these problems, we believe that differences in within-person change across subgroups (or across between-person factors generally) should be assessed and tested on an additive rather than a multiplicative scale. Because standard GEE and GLMM models do not provide such a test, we presented and illustrated a method of analysis which can be used to test whether the rate of change in the probability (rather than the odds) of an event varies across subgroups. Results based on this method were consistent with observed and model-predicted changes in the probability of IADL disability. This was not the case when a logistic product term was used to address the same issue. Analyzing longitudinal change on an additive rather than a multiplicative scale will also make analytic practice with longitudinal data consistent with what has become standard practice in cross-sectional, cohort, and case control studies in epidemiologic research.

While the GEE and GLMM models both indicated that increases in the rate of disability were greater among the cognitively impaired, these rates, and corresponding predicted probabilities were not the same across models. In Table 5, the GEE-based probability of being disabled increased by .07 per year among the intact and by .12 among the impaired. For the GLMM estimates, the corresponding increases were .05 and .27, and the probability of disability at time 3 reached .97 among the impaired. Despite the general preference for the GLMM model in the literature, we find the GEE estimates, which approximate the observed probabilities in Table 2, more plausible, and the findings here lend credence to the concerns of Hubbard et al. (2010) regarding estimates based on the GLMM. While our results raise rather than address this issue, which is beyond the scope of this manuscript, researchers using these methods may want to estimate both models, and report and attempt to explain any discrepant results.

It is useful to distinguish our approach from Norton et al. (2004) who are interested in testing whether statistical interaction is present in the probabilities derived from a logistic model. They show that a logistic model implies a distribution of interactions in the probabilities. Each interaction consists of differences in the marginal effects of a regressor $[= p_i(1-p_i)\beta_{\text{logistic}}]$ at each unique cross-tabulation of all the regressors in one's model. Our substantive interest here is in whether *average* yearly increases in the probability of disability vary across levels of a single between-person factor, and the marginal effects we report are averaged across levels of age, race, gender, and education within levels of baseline impairment. While our approach is consistent with the research question of interest in many longitudinal studies, it does not (as does Norton's approach) test whether statistical interaction is present in the probabilities, and we purposely do not describe our method as a test for interaction. The Norton et al. approach could provide a more complete delineation of how interactive effects vary systematically within as well as between levels of a between-person factor. At present their programs cannot be used with longitudinal data. However, as

shown in (Stata FAQ 2010c), the Stata margins approach can be extended to 3-way and (implicitly) to k-way product terms.

A second limitation of our study is that we do not deal with ‘fixed effects’ approaches to whether the effects of time vary across levels of a between person factor. Compared with the random effects models used here (and in many studies), fixed effects models control more completely for unmeasured covariates and do not require an assumption that these unmeasured factors are uncorrelated with the independent variables in one’s model (Allison 2005). We believe that our study provides evidence that interactions in fixed effects models should also be assessed on an additive scale, but applying our approach to fixed effects models is beyond the scope of the present paper.

Despite these limitations, we believe the methods provided here can be used when researchers have repeated measures data with dichotomous outcome variables and seek to estimate the effects of time constant or time changing independent variables on the direction and rate (trajectory) of change. These might include the effects of such time constant variables as age, gender, and race or the effects of time changing variables such as social support on trajectories of disability. Other empirical studies might examine the effects of these or other predictors on dichotomous outcomes of measures of mental illness, use of medical services, or employment, and so forth.

Before concluding, it is important to note that the problems with a logistic product term identified here apply equally to other ratio-based generalized linear models including Poisson and negative binomial models for counts. In analyses using the latter models with number of ADL disabilities dependent, baseline cognitive impairment was again protective for subsequent increases in disability due to a higher rate of initial disability among the cognitively impaired. A subsequent manuscript will study these problems in detail and offer alternative estimation techniques.

Acknowledgments

This work was supported by the National Institutes of Health, National Institute on Aging, Claude D Pepper Older Americans Independence Center; Grant No.P30 AG028716

References

- Ahlbom, Anders; Alfredsson, Lars. Interaction: A Word with Two Meanings Creates Confusion. *European Journal of Epidemiology*. 2005; 20:563–564. [PubMed: 16119427]
- Allison, Paul D. *A Logistic Regression Using SAS: Theory and Application*. Cary, NC: SAS Institute, Inc; 1999a.
- Allison, Paul D. Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods and Research*. 1999b; 28:186–208.
- Allison, Paul D. *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. Cary, NC: SAS Institute; 2005.
- Assman, Susan F.; Hosmer, David W.; Lemeshow, Stanley; Mundt, Kenneth A. Confidence Intervals for Measures of Interaction. *Epidemiology*. 1996; 7:286–290. [PubMed: 8728443]
- Blazer, Dan; Burchett, Bruce; Service, Connie; George, Linda K. The Association of Age and Depression among the Elderly: An Epidemiologic Exploration. *Journal of Gerontology: Medical Science*. 1991; 46:M210–M215.
- Bollen, Kenneth A. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley; 1989. *Structural Equations with Latent Variables*.
- Buis, Maarten L. Stata tip 87: Interpretation of Interactions in Non-linear Models. *The Stata Journal*. 2010; 10:305–308.

- Cornelißen, Thomas; Sonderhof, Katja. Partial Effects in Probit and Logit Models with a Triple Dummy-Variable Interaction Term. *The Stata Journal*. 2009; 9:571–583.
- Dodge, Hiroko H.; Kadowaki, Takashi; Hayakawa, Takehito; Yamakawa, Masanobu; Sekikawa, Akira; Ueshima, Hirotsugu. Cognitive Impairment as a Strong Predictor of Incident Disability in Specific ADL-IADL Tasks among Community-Dwelling Elders: The Azuchi Study. *Gerontologist*. 2005; 45:222. [PubMed: 15799987]
- Efron, Bradley; Tibshirani, RJ. *An Introduction to the Bootstrap*. New York: Chapman and Hall; 1993.
- Elliott, Delbert S.; Huizinga, David; Menard, Scott. *Multiple Problem Youth: Delinquency, Substance Use, and Mental Health Problems*. New York: SpringerVerlag; 1989.
- Ender, Phil. Margins and the Tao of Interaction. Paper presented at the Boston Stata Conference; July, 2010; 2010.
- Fillenbaum, Gerda G. Screening the Elderly. *Journal of the American Geriatric Society*. 1985; 33:698–706.
- Hanushek, Eric A.; Jackson, John E. *Statistical Methods for Social Scientists*. New York: John Wiley; 1977.
- Hoetker, Glenn. Confounded Coefficients: Extending Recent Advances in the Accurate Comparison of Logit and Probit Coefficients Across Groups. University of Illinois College of Business Working Paper 03–0100. 2004. <http://ssrn.com/abstract=609104> [14 August 2006]
- Hosmer, David W.; Lemeshow, Stanley. Confidence Interval Estimation of Interaction. *Epidemiology*. 1992; 3:452–56. [PubMed: 1391139]
- Hsiao, Cheng. *Analysis of Panel Data*. 2. New York: Cambridge University Press; 2003.
- Hubbard, Alan E.; Jennifer Ahern, Nancy L.; van der Laan Fleischer, Mark; Lippman, Sheri A.; Jewell, Nicholas; Bruckner, Tim; Satariano, William A. To GEE or Not to GEE: Comparing Population Average and Mixed Models for Estimating the Associations Between Neighborhood Risk Factors and Health. *Epidemiology*. 2010; 21:467–474. [PubMed: 20220526]
- Huttenlocher, Janelle; Haight, Wendy; Bryk, Anthony; Seltzer, Michael; Lyons, Thomas. Early Vocabulary Growth: Relations to Language Input and Gender. *Developmental Psychology*. 1991; 27:236–248.
- Kaufman, Jay S. Interaction Reaction. *Epidemiology*. 2009; 20:159–160. [PubMed: 19234409]
- Knol, Mirjam J.; van der Tweel, Ingeborg; Grobbee, Diederick E.; Numans, Mattijs E.; Geerlings, Mirjam I. Estimating Interaction on an Additive Scale between Continuous Determinants in a Logistic Regression Model. *International Journal of Epidemiology*. 2007; 36:1111–1118. [PubMed: 17726040]
- Li, Rongling; Chambless, Lloyd. Test for Additive Interaction in Proportional Hazards Models. *Annals of Epidemiology*. 2007; 17:227–236. [PubMed: 17320789]
- Liang, Kung Yee; Zeger, Scott L. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*. 1986; 73:13–22.
- Loucks, Eric B.; Abrahamowicz, Michael; Xiao, Tongling; Lynch, John W. Associations of Education with 30 Year Life Course Blood Pressure Trajectories: Framingham Offspring Study. *BMC Public Health*. 2011; 10:1186/1471-2458-11-139
- Molenberghs, Geert; Verbeke, Geert. *Models for Discrete Longitudinal Data*. New York: Springer; 2006.
- Moritz, Deborah J.; Kasl, Stanislav V.; Berkman, Lisa F. Cognitive Functioning and the Incidence of Limitations in Activities of Daily Living in an Elderly Community Sample. *American Journal of Epidemiology*. 1995; 141:41–49. [PubMed: 7801965]
- Norton, Edward C.; Wang, Hua; Ai, Chunrong. Computing Interaction Effects and Standard Errors in Logit and Probit Models. *The Stata Journal*. 2004; 4:154–167.
- Palardy, Gregory J.; Rumberger, Russell W. Teacher Effectiveness in First Grade: The Importance of Background Qualifications, Attitudes, and Instructional Practices for Student Learning. *Educational Evaluation and Policy Analysis*. 2008; 30:111–140.
- Pfeiffer E. A Short Portable Mental Status Questionnaire for the Assessment of Organic Brain Deficit in Elderly Patients. *Journal of American Geriatrics Society*. 1975; 23:433–41.

- Pinheiro, José C.; Chao, Edward C. Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models. *Journal of Computational and Graphical Statistics*. 2006; 15:58–81.
- Rabe-Hesketh, Sophia; Skrondal, Anders. *Multilevel and Longitudinal Modeling Using Stata*. Stata Corp; College Station, TX: 2005.
- Raji, Mukaila A.; Ostir, Glenn V.; Markides, Kyriakos S.; Goodwin, James S. 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*. 2002; 57:M678–M682.
- Rodriguez, Germán; Goldman, Noreen. An Assessment of Estimation Procedures for Multilevel Models with Binary Responses. *J Royal Statistical Society, A*. 1995; 158:73–90.
- Rothman, Kenneth J.; Greenland, Sander. *Modern Epidemiology*. Philadelphia, Pa: Lippincott; 1998.
- Rothman, Kenneth J.; Greenland, Sander; Lash, Timothy L. *Modern Epidemiology*. 3. Philadelphia: Lippincott, Williams, and Wilkins; 2008.
- Schafer, Joseph L.; Graham, John W. Missing Data: Our View of the State of the Art. *Psychological Methods*. 2000; 7:147–177. [PubMed: 12090408]
- Shaw BA, Liang Jersey, Krause Neal. Age and Race Differences in the Trajectories of Self-esteem. *Psychology and Aging*. 2010; 25:84–94. [PubMed: 20230130]
- Singer, Judith D.; Willet, John B. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford; NY: 2003.
- SAS Institute Inc. *SAS OnlineDoc 9.1.3*. Cary, NC: SAS Institute Inc; 2005.
- StataCorp. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP; 2009.
- Stata FAQ. Statistical Computing Seminars: Deciphering Interactions in Logistic Regression. Vol. 2010a. UCLA: Academic Technology Services, Statistical Consulting Group; 2010a. Retrieved January 2011. www.ats.ucla.edu/stat/stata/seminars/interaction_sem/interaction_sem.htm
- Stata FAQ. How Can I Use the Margins Command to Understand Multiple Interactions in Logistic Regression?. UCLA: Academic Technology Services, Statistical Consulting Group; 2010b. 2010b. Retrieved January 2011. http://www.ats.ucla.edu/stat/stata/faq/margins_mlogcatcon.htm
- Stata FAQ. Parts 1–2: “How Can I Understand a 3-way Continuous Interaction?”. UCLA: Academic Technology Services, Statistical Consulting Group; 2010c. Retrieved November 24, 2010. <http://www.ats.ucla.edu/stat/stata/faq/con3way11.htm>
- Twisk, Joseph W. Longitudinal Data Analysis. A Comparison Between Generalized Estimating Equations and Random Coefficient Analysis. *European Journal of Epidemiology*. 2004; 19:769–776. [PubMed: 15469034]
- VanderWeele, Tyler J.; Robins, James M. Empirical and Counterfactual Conditions for Sufficient Cause Interactions. *Biometrik*. 2008; 95:49–61.
- VanderWeele, Tyler J. Sufficient Cause Interactions and Statistical Interactions. *Epidemiology*. 2009; 20:6–13. [PubMed: 19234396]
- Walls, Theodore A.; Shafer, Joseph L. *Models for Intensive Longitudinal Data*. New York: Oxford University Press; 2006.
- Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press; 2002.

Highlights

- With repeated measures models, logistic interactive terms can be misleading
- Differences in the rate of change should be assessed on an additive scale
- This can be done using the STATA margins command

Changes in the Proportion Disabled by Cognitive Impairment Over 4 (Yearly) Waves of EPESE Data(N=4066).

Table 1

	Time 0	Time 1	Time 2	Time 3	Mean yearly change
Cognitively intact:	.04 (3528)	.09 (2846)	.17 (2225)	.23 (1516)	+.063
Cognitively impaired:	.36 (538)	.47 (348)	.53 (199)	.61 (96)	+.083
	(4066)	(3194)	(2454)	(1612)	

Note: Respondents were asked whether they could perform each of 5 instrumental ADL tasks (traveling, shopping, preparing meals, doing housework, managing finances without assistance (Fillenbaum, G.G., 1985). Respondents unable to perform 4 or 5 tasks were classified as disabled.

Changes in the Probability and the Odds of IADL Disability by Cognitive Impairment Over 2 Yearly Waves of EPESE Data (N=4066)

Table 2

<u>No Cognitive Impairment</u>			<u>Cognitive Impairment</u>				
	Prob.	Odds	N	Prob.	Odds	N	
Time 0:	.044	.046	3528	Time 0:	.355	.550	538
Time 1:	.087	.095	2846	Time 1:	.466	.873	348
Additive scale: $\Delta Prob/\Delta t = .087-.044 = .04$				Additive scale: $\Delta Prob/\Delta t = .466-.355 = .11$			
Multiplicative scale: $\Delta Odds/\Delta t = .095/.046 = 2.07$				Multiplicative scale: $\Delta Odds/\Delta t = .873/.550 = 1.59$			
<hr/>							
Product term:							
<i>Multiplicative scale: 1.59/2.07 = .77, p=.13</i>							
<i>Additive scale: .11-.04 = .07, p=.01;</i>							
<hr/>							

Table 3
The Impact of Baseline Prevalence (BP) on Additive and Multiplicative Product Terms (Hypothetical Data)

	Low Risk Group BP=.05		High Risk Group(1) BP=.05		High Risk Group(2) BP=.10		High Risk Group(3) BP=.30	
	P	ODDS	P	ODDS	P	ODDS	P	ODDS
Time 0	.05	.053	.05	.053	.10	.111	.30	.429
Time 1	.10	.111	.15	.176	.20	.250	.40	.667
$\Delta Prob/\Delta t^a$	.05		.10		.10		.10	
$\Delta Odds/\Delta t^b$	2.09		3.32		2.25		1.55	
Product term:								
Additive			.10 – .05 = .05		.10 – .05 = .05		.10 – .05 = .05	
Multiplicative			3.32/2.09= 1.59		2.25/2.09 = 1.08		1.55/2.09 = .74	

Notes:

^aChange on an additive scale = (probability of disability at time 1)-(probability of disability at baseline).

^bChange on a multiplicative scale = (odds of disability at time 1)/(odds of disability at baseline).

GEE and GLMM Results for Disability by Time by Cognitive Impairment Over Four Waves of EPESE data.

Table 4

Fixed Effects:	GEE			GLMM		
	B	(SE)	OR	B	(SE)	OR
Intercept	-2.991	(.065)	.050**	-7.205	(.372)	0.001**
Cognitive Impairment	2.421	(.109)	11.257**	5.604	(.384)	270.400**
Cog. Impairment X Time	-0.198	(.068)	.820**	0.034	(.140)	1.034
Time among intact	0.698	(.028)	2.009**	1.677	(.107)	5.349**
Time among impaired	0.500	(.063)	1.648**	1.711	(.152)	5.534**
Random effects:				EST.	(SE)	SIG.
Intercept variance				15.017	(1.757)	**
Predicted Probabilities:	INTACT	IMPAIRED	INTACT	IMPAIRED		
Time 0	.048	.361	.001	.168		
Time 1	.092	.482	.004	.528		
Time 2	.169	.606	.021	.861		
Time 3	.290	.717	.102	.972		
$\Delta Prob/\Delta t$: mean yearly increase:	.081	.119	.034	.268		

Note:

* indicates p<.05;
** indicates p<.01;

Table 5

GEE and GLMM Results with Marginal Effects for Impairment by Time Interaction Controlling for Age, Race, Gender, and Education; Four Waves of EPESE data.

	GEE			GLMM		
	B	(SE)	OR	B	(SE)	OR
<u>Fixed Effects:^a</u>						
Intercept	-11.150	(.454)	.00001**	-23.30	(1.328)	7.633-e11**
Cognitive Impairment	1.889	(.091)	6.61**	3.381	(.387)	29.40**
Time (all respondents)	.799 ^b	(.030)	2.22**			
Cog. Impairment X Time				.571	(.245)	1.770**
Time (intact)				1.650	(.085)	5.206**
(impaired)				2.221	(.228)	9.217**
<u>Random effects:</u>				EST.	(SE)	SIG.
Intercept variance (intact)				9.41	(1.338)	**
Intercept variance (impaired)				20.81	(4.369)	**
<u>Marginal effects:</u>	dy/dt	(SE)	SIG.	dy/dt	(SE)	SIG.
cog. intact	.072	(.003)	**	.054	(.003)	**
cog. impaired	.158	(.005)	**	.269	(.016)	**
group difference	.086	(.002)	**	.214	(.016)	**
<u>Predicted Probabilities:</u>	INTACT	IMPAIRED	IMPAIRED	INTACT	IMPAIRED	IMPAIRED
Time 0	.048	.344		.007	.230	
Time 1	.093	.470		.026	.472	
Time 2	.164	.598		.070	.752	
Time 3	.261	.715		.166	.936	
$\Delta Prob/\Delta t$: mean yearly increase:	.071	.124		.053	.235	

Notes:

* indicates $p < .05$;

** indicates $p < .01$;

^a Covariates are age in years, African-American vs. other, gender, and education in years.

^b With covariates, cognitive impairment by time term was not significant and dropped from the GEE model.