

SC708: Hierarchical Linear Modeling
Instructor: Natasha Sarkisian
Class notes: Non-continuous Dependent Variables

So far we've only dealt with continuous dependent variables, but HLM allows us to estimate models when the dependent variables are categorical. Your dependent variable can be dichotomous (0/1), categorical with multiple unordered categories, ordinal, or count variable. In such cases, linear models are inappropriate as there are no restrictions on the predicted values of level-1 outcome, the level-1 random effect (i.e. level 1 residual) cannot be normally distributed, and cannot have homogenous variance (the variance depends on the predicted value). Therefore, we need to use HGLM models for such variables. Like in non-hierarchical analysis, this is accomplished by specifying a link function that transforms the dependent variable so that the level-1 predicted values are constrained to be within a specific interval. Specifically, we use logit models for dichotomous variables, multinomial logit for categorical with unordered categories, ordered logit for ordinal variables, and Poisson models for count variables.

The following is an example of analysis with a dichotomous dependent variable. We'll use THAIUGRP.MDM in Examples/Chapter 6. These are data on 7,516 sixth graders nested within 356 primary schools from a national survey of primary education in Thailand, conducted in 1988. The dependent variable of interest is the probability that a child will repeat a grade during the primary years (REP1). The level-1 independent variables are whether a child attended pre-primary education (PPED) and child's gender (MALE). The level-2 variable is mean SES of school (MSESC). To specify that the dependent variable is binary, we go to the Basic Settings menu and select Bernoulli option.

LEVEL 1 MODEL

$$\text{Prob}(\text{REP1}_{ij}=1|\beta_j) = \varphi_{ij}$$

$$\text{Log}[\varphi_{ij}/(1 - \varphi_{ij})] = \eta_{ij}$$

$$\eta_{ij} = \beta_{0j} + \beta_{1j}(\text{MALE}_{ij}) + \beta_{2j}(\text{PPED}_{ij})$$

LEVEL 2 MODEL

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{MSESC}_j - \overline{\text{MSESC}_.}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{MSESC}_j - \overline{\text{MSESC}_.}) + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{MSESC}_j - \overline{\text{MSESC}_.}) + u_{2j}$$

The model specified for the fixed effects was:

```
-----
Level-1                               Level-2
Coefficients                           Predictors
-----
      INTRCPT1, B0                      INTRCPT2, G00
$                                     MSESC, G01
      MALE slope, B1                    INTRCPT2, G10
$                                     MSESC, G11
      PPED slope, B2                   INTRCPT2, G20
$                                     MSESC, G21
'$' - This level-2 predictor has been centered around its grand mean.
```

Summary of the model specified (in equation format)

Level-1 Model

Fixed Effect	Coefficient	Error	T-ratio	d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	-2.043050	0.094461	-21.628	354	0.000
MSESC, G01	-0.410774	0.255098	-1.610	354	0.108
For MALE slope, B1					
INTRCPT2, G10	0.465559	0.075695	6.150	354	0.000
MSESC, G11	0.270760	0.204300	1.325	354	0.186
For PPED slope, B2					
INTRCPT2, G20	-0.532227	0.095718	-5.560	354	0.000
MSESC, G21	-0.044859	0.253801	-0.177	354	0.860

Fixed Effect	Coefficient	Odds Ratio	Confidence Interval
For INTRCPT1, B0			
INTRCPT2, G00	-2.043050	0.129633	(0.108,0.156)
MSESC, G01	-0.410774	0.663137	(0.402,1.094)
For MALE slope, B1			
INTRCPT2, G10	0.465559	1.592905	(1.373,1.848)
MSESC, G11	0.270760	1.310960	(0.878,1.958)
For PPED slope, B2			
INTRCPT2, G20	-0.532227	0.587296	(0.487,0.709)
MSESC, G21	-0.044859	0.956132	(0.581,1.574)

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	1.14692	1.31543	237	423.83711	0.000
MALE slope, U1	0.31659	0.10023	237	213.85874	>.500
PPED slope, U2	0.30018	0.09011	237	166.92266	>.500

Final estimation of fixed effects: (Population-average model)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	-1.671706	0.082783	-20.194	354	0.000
MSESC, G01	-0.398163	0.221047	-1.801	354	0.072
For MALE slope, B1					
INTRCPT2, G10	0.426803	0.060136	7.097	354	0.000
MSESC, G11	0.248359	0.162132	1.532	354	0.126
For PPED slope, B2					
INTRCPT2, G20	-0.481299	0.079160	-6.080	354	0.000
MSESC, G21	-0.000757	0.215165	-0.004	354	0.997

Fixed Effect	Coefficient	Odds Ratio	Confidence Interval
For INTRCPT1, B0			
INTRCPT2, G00	-1.671706	0.187926	(0.160,0.221)
MSESC, G01	-0.398163	0.671553	(0.435,1.037)
For MALE slope, B1			
INTRCPT2, G10	0.426803	1.532350	(1.362,1.724)
MSESC, G11	0.248359	1.281920	(0.932,1.763)
For PPED slope, B2			
INTRCPT2, G20	-0.481299	0.617980	(0.529,0.722)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
MSESC, G21	-0.000757	0.999243		(0.655, 1.525)	

Final estimation of fixed effects (Population-average model with robust standard errors)					

For INTRCPT1, B0					
INTRCPT2, G00	-1.671706	0.059774	-27.967	354	0.000
MSESC, G01	-0.398163	0.170459	-2.336	354	0.020
For MALE slope, B1					
INTRCPT2, G10	0.426803	0.045573	9.365	354	0.000
MSESC, G11	0.248359	0.133309	1.863	354	0.063
For PPED slope, B2					
INTRCPT2, G20	-0.481299	0.059564	-8.080	354	0.000
MSESC, G21	-0.000757	0.173800	-0.004	354	0.997

Fixed Effect	Coefficient	Odds Ratio	Confidence Interval		

For INTRCPT1, B0					
INTRCPT2, G00	-1.671706	0.187926	(0.167, 0.211)		
MSESC, G01	-0.398163	0.671553	(0.480, 0.939)		
For MALE slope, B1					
INTRCPT2, G10	0.426803	1.532350	(1.401, 1.676)		
MSESC, G11	0.248359	1.281920	(0.987, 1.666)		
For PPED slope, B2					
INTRCPT2, G20	-0.481299	0.617980	(0.550, 0.695)		
MSESC, G21	-0.000757	0.999243	(0.710, 1.406)		

Interpreting fixed effects

The interpretation of the fixed effects is very similar to the interpretation of the results of logistic regression—but be careful as we now have variables on multiple levels and can potentially have interactions across levels.

Interpreting coefficients themselves allows us to discuss the direction and significance of effects, but not their size. To talk about the size, we use odds ratios.

Odds are ratios of two probabilities – probability of a positive outcome and a probability of a negative outcome (e.g. probability of voting divided by a probability of not voting). But since probabilities vary depending on values of X, such a ratio varies as well. What remains constant is the ratio of such odds – e.g. odds of repeating a grade for male child divided by odds of repeating a grade for a female child will be the same number regardless of the values of other variables on the model. Similarly, the odds ratio for age can be a ratio of the odds of repeating a grade for someone who is 12 y.o. to the odds of an 11 y.o. child, or of a 17 y.o. to a 16 y.o. child's odds – these will be the same regardless of what age values you pick, as long as they are one year apart.

Odds ratios are exponentiated logistic regression coefficients. They are sometimes called factor coefficients, because they are multiplicative coefficients. Odds ratios are equal to 1 if there is no effect, smaller than 1 if the effect is negative and larger than 1 if it is positive. So for example, the odds ratio for 1.53 for males indicates that the odds of repeating a grade for males are 1.53

times higher than for females –or we can say 53% higher. And the odds ratio of 0.67 for MSES indicates that each additional unit increase in mean SES of school is associated in a 33% decrease in a student's odds of repeating a grade. To get percent change, we subtract 1 from the odds ratio, and then multiply the result by 100.

Beware: if you would like to know what the increase would be per, say, 10 units increase in the independent variable – e.g. 10 years of age or education, you cannot simply multiply the odds ratio by 10! The coefficient, in fact, would be odds ratio to the power of 10. Or alternatively, you could take the regular logit coefficient, multiply it by 10 and then exponentiate it.

In addition, since odds ratios are multiplicative coefficients, when you want to interpret, for example, an interaction term, you would have to multiply rather than add the odds ratio numbers. Alternatively, you can add the numbers presented in the coefficient column and then exponentiate the result.

In addition to using odds ratios, we can use predicted probabilities (P) to interpret our results. HLM does not calculate predicted probabilities but we can get them by calculating predicted logits (L) and then recalculating them into probabilities. Since $L = \log(\text{odds}) = \log(P/(1-P))$, then $P = e^L / (1 + e^L)$.

As mentioned above, predicted logits L are available in the FITVAL variable, so you can easily generate predicted probabilities on the basis of that. It is more interesting for interpretation purposes, however, to calculate predicted probabilities for some hypothetical, strategically selected cases. For that, you have to calculate the logit of interest by hand by plugging values into the equation:

$$L = \gamma_{00} + \gamma_{01} * \text{MSES} + \gamma_{10} * \text{MALE} + \gamma_{11} * \text{MALE} * \text{MSES} + \gamma_{20} * \text{PPED} + \gamma_{21} * \text{PPED} * \text{MSES}$$

For instance, if we want to calculate the predicted probabilities for males and females who did not attend pre-primary education and who are in a school with average SES, we calculate:

$$L \text{ for males} = \gamma_{00} + \gamma_{10} * \text{MALE} = -1.67 + 0.43 = -1.24$$

$$L \text{ for females} = \gamma_{00} = -1.67$$

Corresponding probabilities would be:

$$P \text{ for males} = \exp(-1.24) / (1 + \exp(-1.24)) = .22$$

$$P \text{ for females} = \exp(-1.67) / (1 + \exp(-1.67)) = .16$$

Such strategically calculated predicted probabilities are very useful for a more intuitive presentation of results. Note, however, that the differences between probabilities are not constant – e.g. if we looked at those who attended preprimary education (PPED=1), the difference between males and females would change.

We can also calculate predicted probabilities for specific schools – we calculate linear predictors and add the corresponding level 2 residual. For example, we take female students with no preprimary education in schools 10103 and 10104. We would need to know MSES and EBINTRCPT1 for their schools. Since we used MSES as a mean-centered variable, we need to

get its values based on that; for that, we generate a mean-centered variable in our level 2 residuals file:

```
. sum MSESC
      Variable |      Obs      Mean   Std. Dev.   Min      Max
-----+-----
      MSESC |      356   .0078371   .3806623   -.77     1.49
. gen  MSESCm = MSESC-.0078371
```

School	N	MSESCm	L for MALE=0 PPED=0 ($L=\gamma_{00} + \gamma_{01} * MSESC$)	Predicted probability based on fixed effects only	Random effect EBINTRCPT1	Predicted probability based on fixed and random effects
10103	17	.87	-1.67-0.4*.87=-2.018	.117	-.216	.097
10104	29	.19	-1.67-0.4*.19=-1.746	.149	-1.228	.049

Thus we see that school 10104 has advantages because of its mean SES, but its unique component suggests that it is much less efficient than school 10104 – in fact the unique component of school 10104 more than compensates for its economic disadvantage.

Note that if we focused on male students, we’d have to take into account both the fixed effect (coefficient) for MALE and the value of random effect for the MALE slope which is stored in EBMALE.

Interactions

Note that interactions as a method to compare two or more groups can be problematic in logit models because the coefficients are scaled according to the differences in residual dispersion. So it is not as appropriate to rely on the significance test of the interaction term to establish whether some process differs by group. This is especially problematic for HLM as it heavily relies on the use of cross-level interactions. The best approach to establish whether the two groups do differ is to examine differences in predicted probabilities. You would have to do that in Stata, and you would have to decide which values to assign to the rest of the variables in your model.

```
. sum msesc
      Variable |      Obs      Mean   Std. Dev.   Min      Max
-----+-----
      msesc |     7516   .009674   .3759242   -.77     1.49

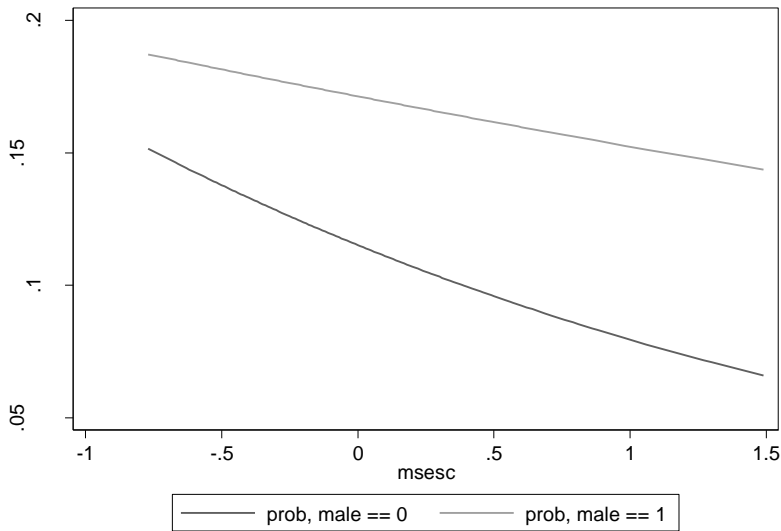
. gen msescm=msesc-r(mean)

. gen pred= -2.043050 - 0.410774*msescm + 0.465559*male + 0.270760*male*msescm-
0.532227*0 - 0.044859*0*msescm

. gen prob=exp(pred)/(1+exp(pred))

. separate prob, by(male)
      storage display   value
variable name  type  format  label  variable label
-----+-----
prob0          float  %9.0g                prob, male == 0
prob1          float  %9.0g                prob, male == 1
```

```
. line prob0 prob1 msesc, sort
```



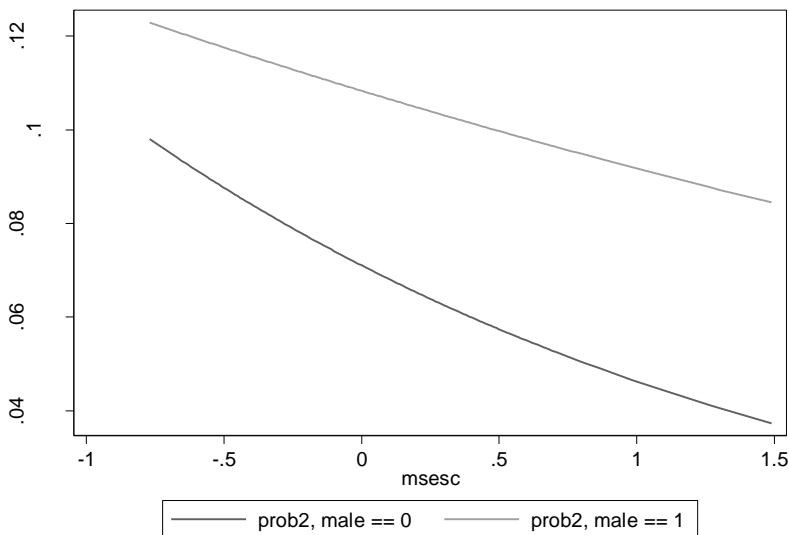
```
. gen pred2= -2.043050 - 0.410774*msescm + 0.465559*male + 0.270760*male*msescm-
0.532227*1 - 0.044859*1*msescm
```

```
. gen prob2=exp(pred2) / (1+exp(pred2))
```

```
. separate prob2, by(male)
```

variable name	storage type	display format	value label	variable label
prob20	float	%9.0g		prob2, male == 0
prob21	float	%9.0g		prob2, male == 1

```
. line prob20 prob21 msesc, sort
```



For more detail and more complex ways to do these comparisons, see Scott Long's article at: http://www.indiana.edu/~jslsoc/files_research/groupdif/groupwithprobabilities/groups-with-prob-2009-06-25.pdf

Variance components

Note that the variance component does not contain an estimate of level 1 variance. That is because in logistic regression models, it is not possible to estimate both the coefficients and the error variance; therefore, in all logistic regression models, the error variance is always fixed to the same number which is $\pi^2/3 = 3.29$. That rule also applies to multilevel models, but only to their level 1 residuals. Knowing this means that we can calculate the intraclass correlation coefficient or the proportion of variance explained. For both, we can follow the procedures described on pp.224-227 of the Snijders and Bosker chapter on dichotomous outcomes. For instance, the ICC would be calculated as

$$\rho_i = \frac{\tau_0^2}{\tau_0^2 + \pi^2/3}$$

And the proportion of variance explained can be calculated as

$$R_{\text{dicho}}^2 = \frac{\sigma_F^2}{\sigma_F^2 + \tau_0^2 + \sigma_R^2}$$

Note that in addition to the level 2 intercept variance τ_0 and level 1 variance $\sigma_R^2 = 3.29$, we need to know the variance of fitted values σ_F^2 . That refers to the variance of linear predictions, which are the values that results if we multiply our coefficients by our variable values and add up these products. That is, we are talking about the predicted values of logits. To obtain the variance of fitted values, we can use level 1 residuals file and calculate the variance of the FITVAL variable containing the linear predictor values. Note that such R squared values are typically lower than values we are used to with OLS because σ_R^2 is a fixed number.

Unit-specific versus population-average models

The distinction between unit-specific and population-average models emerges when we use nonlinear link models (HGLM). The unit-specific model presents coefficients for a hypothetical unit (group) where random effect is zero. The population-average model presents coefficients averaged out for the whole sample.

Further, these two models make different assumptions about the underlying distribution of random effects and they are oriented towards different research aims. The unit-specific models are more appropriate for describing how the effects of level 1 predictors vary across level 2 units. Population-average models, in contrast, give answers to population-average questions – that is, they are more appropriate for estimating predicted probabilities for the whole population. If we use a regression model to examine how preprimary education experience relates to the risk of class repetition in different schools, we are asking a unit-specific question. If we want to know how the risk of repetition differs between those who do and do not have preprimary experience nationwide, we need a population-average estimate.

Also note that population-average inferences are based on fewer assumptions than unit-specific inferences and are therefore more robust to erroneous assumptions about the random effects in the model. In a way, unit-specific models are richer but more sensitive to model assumptions.

In regular HLM, these are the same, but in HGLM they differ because nonlinear transformations, such as that from probability into log odds, mean that the distribution of predicted probabilities is not symmetric. The two estimates become rather similar when the fixed effect is close to 0 or the random component is close to 0.

Ordered logit and multinomial logit

Much of what we discussed applies to ordered logit and multinomial logit models. Note that when specifying either ordered or multinomial logit, you will have to indicate the number of categories of your DV in HLM. To better understand interpretation of coefficients in ordered logit and multinomial logit, you should review my SC704 notes for each of these topics.

Briefly, the odds ratios for ordered logit are cumulative odds of belonging to a certain category or lower versus belonging to one of the higher categories. For example, if our dependent variable is the level of agreement with some statement and the categories are agree=3, not sure=2, and disagree=1, and if the odds ratio for gender as a predictor of that agreement is 2.00, we can say that the odds of disagreeing rather than agreeing or being not sure are 2 times higher for women than for men. Similarly, the odds of disagreeing or being not sure are also twice as high for women than for men. What this means is that ologit assumes that these two odds ratios are essentially the same and thus uses the average. That is called the parallel slopes assumption. So we are assuming these two odds ratios are the same – if they differ significantly, the assumption is violated.

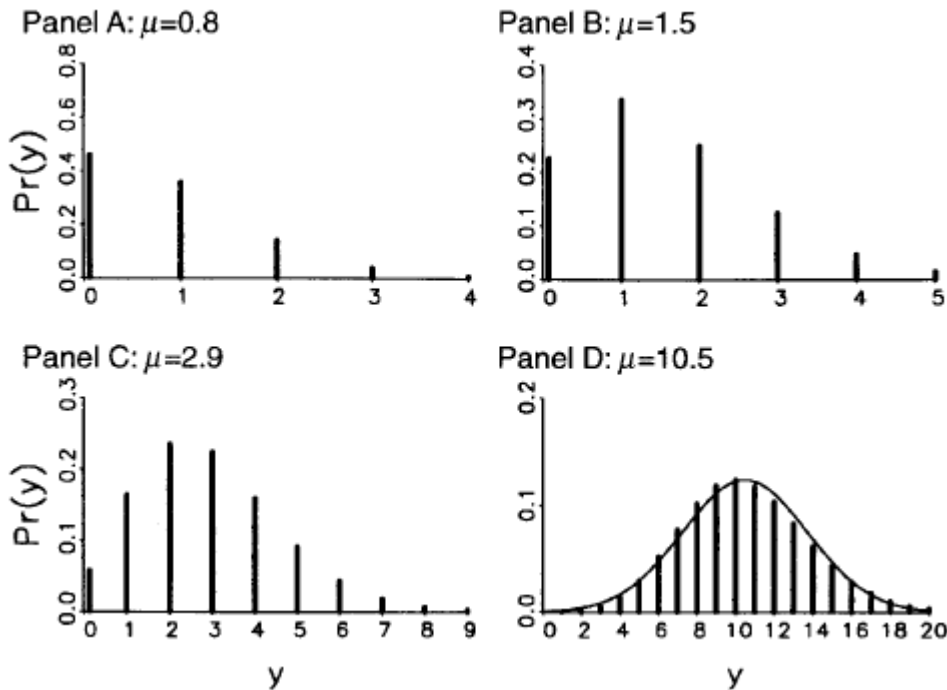
Also, you should be aware that HLM does not provide diagnostic tools for testing such assumptions specific to ordered logit (parallel slopes assumption) or multinomial logit (independence of irrelevant alternatives), so in order to obtain a rough test of those, you might want to run your model in Stata and test these assumptions there.

What we should also note that ordered logit coefficients in HLM go in the opposite direction to what ordered logit coefficients are in Stata: They show how the odds of being in a lower category change, while standard ordered logit coefficients show how the odds of being in a higher category change. Hence they would be the same number but with an opposite sign.

Count data models

HLM also has the capability for count data models. Count variables are often treated as though they are continuous, and regular regression is used, but it can result in inefficient, inconsistent, and biased estimates. Need to use models that are developed specifically for count data. Poisson model is the most basic of them.

Poisson distributions:



Characteristics of Poisson distribution:

1. $E(y) = \mu$
2. The variance equals the mean: $\text{Var}(y)=E(y)= \mu$ -- equidispersion. In practice, the variance is often larger than μ : this is called overdispersion. The main reason for overdispersion is heterogeneity – if there are different groups within data that have different means and all of them are actually equal to their variances, when you put all of these groups together, the resulting combination will have variance larger than the mean. Therefore, we need to control for all those sources of heterogeneity. Thus, when using Poisson regression, we need to ensure that the conditional variance equals to the mean – that is $\text{Var}(y|X)=E(y|X)$.
3. As μ increases, the probability of zeros decreases. But for many count variables, there are more observed zeros than would be predicted from Poisson distribution
4. As μ increases, the Poisson distribution approximates normal.
5. The assumption of independence of events – past outcomes don't affect future outcomes.

Luckily, in HLM we can estimate Poisson model, both with and without overdispersion. Unfortunately, HLM does not provide the output for the overdispersion parameter, so there is no significance test to establish whether it is needed. But, you can compare the results with and without and if the findings are different, go with the overdispersion model.

In addition, HLM models for count data also allow controlling for so-called exposure – that is usually a variable that indicates how long there has been an opportunity to accumulate counts. For example, if we have a count of missed classes from students in different schools, but different schools have different number of days in their school year, then some students have more opportunity to miss classes than others and we need to adjust for exposure – i.e. rather than examine the total count, we would examine the number of missed classes per school day.

Let's examine an example of count data model in HLM and briefly discuss interpretation.

We will use another dataset based on the same national survey of primary education in Thailand in 1988; the file is THAIGRP.MDM in Chapter 6. These are data on 1097 children repeating a grade during their time at primary school where REP1 is the number of grade retentions for each of four subpopulations within a specific school created based on combinations of predictors MALE (1= male, 0 = female) and preschool experience PPED (1= yes, 0= no). So we have a maximum of four observations per school in this dataset. TRIAL is the number of students in a specific subpopulation within a specific school. Finally, MSESC is the average socio-economic status score.

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

$$E(\text{REP1}|\beta) = \lambda$$

$$\text{Log}[\lambda] = \eta$$

$$\eta = \beta_0 + \beta_1(\text{MALE}) + \beta_2(\text{PPED})$$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$$\beta_0 = \gamma_{00} + \gamma_{01}(\text{MSESC}) + u_0$$

$$\beta_1 = \gamma_{10} + \gamma_{11}(\text{MSESC}) + u_1$$

$$\beta_2 = \gamma_{20} + \gamma_{21}(\text{MSESC}) + u_2$$

Distribution at Level-1: Poisson

Weighting Specification

	Weighting?	Weight Variable Name	Normalized?
Level 1	no		
Level 2	no		
Precision	no		

The outcome variable is REP1

The model specified for the fixed effects was:

	Level-1 Coefficients	Level-2 Predictors
	INTRCPT1, B0	INTRCPT2, G00
\$		MSESC, G01
	MALE slope, B1	INTRCPT2, G10
\$		MSESC, G11
	PPED slope, B2	INTRCPT2, G20
\$		MSESC, G21

'\$' - This level-2 predictor has been centered around its grand mean.

The model specified for the covariance components was:

Tau dimensions
INTRCPT1

MALE slope
 PPED slope

Summary of the model specified (in equation format)

 Level-1 Model

$$E(Y|B) = L$$

$$V(Y|B) = L$$

$$\log[L] = B0 + B1*(MALE) + B2*(PPED)$$

Level-2 Model

$$B0 = G00 + G01*(MSESC) + U0$$

$$B1 = G10 + G11*(MSESC) + U1$$

$$B2 = G20 + G21*(MSESC) + U2$$

Level-1 variance = 1/L

Note: The chi-square statistics reported above are based on only 239 of 356 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

RESULTS FOR NON-LINEAR MODEL WITH THE LOG LINK FUNCTION: Unit-Specific Model
 (macro iteration 255)

Tau			
INTRCPT1,B0	0.82061	-0.00557	-0.60413
MALE,B1	-0.00557	0.12963	-0.01266
PPED,B2	-0.60413	-0.01266	1.11767

Tau (as correlations)

INTRCPT1,B0	1.000	-0.017	-0.631
MALE,B1	-0.017	1.000	-0.033
PPED,B2	-0.631	-0.033	1.000

Random level-1 coefficient	Reliability estimate

INTRCPT1, B0	0.373
MALE, B1	0.081
PPED, B2	0.364

Note: The reliability estimates reported above are based on only 239 of 356 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

The value of the likelihood function at iteration 2 = -1.708667E+003

The outcome variable is REPI

Final estimation of fixed effects: (Unit-specific model)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value	

For INTRCPT1, B0						
INTRCPT2, G00	-0.398269	0.079559	-5.006	354	0.000	
MSESC, G01	-0.657354	0.210343	-3.125	354	0.002	
For MALE slope, B1						
INTRCPT2, G10	0.345354	0.068031	5.076	354	0.000	

	MSESC, G11	0.154681	0.177184	0.873	354	0.384
For	PPED slope, B2					
	INTRCPT2, G20	-0.258831	0.098413	-2.630	354	0.009
	MSESC, G21	0.503835	0.261474	1.927	354	0.054

Fixed Effect	Coefficient	Event Rate Ratio	Confidence Interval
For	INTRCPT1, B0		
	INTRCPT2, G00	-0.398269	(0.574,0.785)
	MSESC, G01	-0.657354	(0.343,0.783)
For	MALE slope, B1		
	INTRCPT2, G10	0.345354	(1.236,1.614)
	MSESC, G11	0.154681	(0.824,1.653)
For	PPED slope, B2		
	INTRCPT2, G20	-0.258831	(0.636,0.937)
	MSESC, G21	0.503835	(0.990,2.766)

The outcome variable is REPl

Final estimation of fixed effects
(Unit-specific model with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value	
For	INTRCPT1, B0					
	INTRCPT2, G00	-0.398269	0.079919	-4.983	354	0.000
	MSESC, G01	-0.657354	0.227792	-2.886	354	0.005
For	MALE slope, B1					
	INTRCPT2, G10	0.345354	0.067476	5.118	354	0.000
	MSESC, G11	0.154681	0.172418	0.897	354	0.371
For	PPED slope, B2					
	INTRCPT2, G20	-0.258831	0.099240	-2.608	354	0.010
	MSESC, G21	0.503835	0.273354	1.843	354	0.066

Fixed Effect	Coefficient	Event Rate Ratio	Confidence Interval
For	INTRCPT1, B0		
	INTRCPT2, G00	-0.398269	(0.574,0.786)
	MSESC, G01	-0.657354	(0.331,0.811)
For	MALE slope, B1		
	INTRCPT2, G10	0.345354	(1.237,1.613)
	MSESC, G11	0.154681	(0.832,1.638)
For	PPED slope, B2		
	INTRCPT2, G20	-0.258831	(0.635,0.938)
	MSESC, G21	0.503835	(0.968,2.831)

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	0.90588	0.82061	237	379.71671	0.000
MALE slope, U1	0.36004	0.12963	237	221.96128	>.500
PPED slope, U2	1.05720	1.11767	237	305.04990	0.002

Note: The chi-square statistics reported above are based on only 239 of 356 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

RESULTS FOR NON-LINEAR MODEL WITH THE LOG LINK FUNCTION:
Population Average Model

The value of the likelihood function at iteration 2 = -1.843090E+003

The outcome variable is REPl

Final estimation of fixed effects: (Population-average model)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	-0.135548	0.072966	-1.858	354	0.064
MSESC, G01	-0.649303	0.200682	-3.235	354	0.002
For MALE slope, B1					
INTRCPT2, G10	0.395237	0.062971	6.277	354	0.000
MSESC, G11	0.159484	0.173161	0.921	354	0.358
For PPED slope, B2					
INTRCPT2, G20	-0.307032	0.087766	-3.498	354	0.001
MSESC, G21	0.438226	0.240192	1.824	354	0.068

Fixed Effect	Coefficient	Event Rate Ratio	Confidence Interval
For INTRCPT1, B0			
INTRCPT2, G00	-0.135548	0.873238	(0.757,1.008)
MSESC, G01	-0.649303	0.522410	(0.352,0.775)
For MALE slope, B1			
INTRCPT2, G10	0.395237	1.484736	(1.312,1.680)
MSESC, G11	0.159484	1.172905	(0.835,1.648)
For PPED slope, B2			
INTRCPT2, G20	-0.307032	0.735627	(0.619,0.874)
MSESC, G21	0.438226	1.549955	(0.967,2.484)

The outcome variable is REPl

Final estimation of fixed effects
(Population-average model with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	-0.135548	0.072503	-1.870	354	0.062
MSESC, G01	-0.649303	0.240099	-2.704	354	0.008
For MALE slope, B1					
INTRCPT2, G10	0.395237	0.059300	6.665	354	0.000
MSESC, G11	0.159484	0.170298	0.936	354	0.350
For PPED slope, B2					
INTRCPT2, G20	-0.307032	0.091403	-3.359	354	0.001
MSESC, G21	0.438226	0.284485	1.540	354	0.124

Fixed Effect	Coefficient	Event Rate Ratio	Confidence Interval
For INTRCPT1, B0			
INTRCPT2, G00	-0.135548	0.873238	(0.757,1.007)
MSESC, G01	-0.649303	0.522410	(0.326,0.837)
For MALE slope, B1			
INTRCPT2, G10	0.395237	1.484736	(1.322,1.668)
MSESC, G11	0.159484	1.172905	(0.839,1.639)
For PPED slope, B2			
INTRCPT2, G20	-0.307032	0.735627	(0.615,0.880)
MSESC, G21	0.438226	1.549955	(0.886,2.710)

With regular coefficients, we can interpret sign and significance, but to interpret the size, we exponentiate the coefficients – these are called event rate ratios in HLM or sometimes they are also called incidence-rate ratios. They are also multiplicative coefficients, like odds ratios, and can be interpreted as percent change in the number of events.

Overall, some of the same concerns apply here – those in terms of interactions, in terms of variance. The level 1 residuals variance, which we assumed to be 3.29 in logit-based models, is assumed to be equal to predicted mean, so you need to find the value of average predicted count by generating predicted values, exponentiating them, and calculating their mean. You can use that number as level 1 variance in the formulas discussed above.