

Hierarchical Linear Modeling and Latent Growth Analysis In Public Policy Research

Philip W. Wirtz
The George Washington University

GWIPP Workshop, March 4, 2004

Ecological Fallacy

Robinson (1950):

- **From 1930 Census data, two measures in each of 48 states:**
 - **Percent of the population who are foreign born (“nativity”)**
 - **Percent of the population who are literate**
- **Noted a highly positive correlation (0.53), suggesting that the foreign born are more likely to be literate than the native born.**

- **In fact, exactly the opposite is correct: correlation computed at the individual level is -0.12 .**

- **Ecological correlation gives incorrect inference**
 - **The sign of the correlation is positive because the foreign born tended to live in states where the nativeborn are relatively literate.**

Drives home the importance of analyzing at the proper level.

The Baby and the Bathwater

- **There might still be an interest in statewide measures**

The relationship between nativity and literacy might differ across states, as a function of (for example) the percentage of the state's inhabitants who are foreign born

This implies the need for a multi-level (or “nested” model), where measures at both the individual and state level figure into the analysis.
- **The Problem**

Until relatively recently, a serious methodological limitation precluded us from entering variables from multiple levels into a single model. We were “stuck” with using variables from a single level.
- **The Solution: Dempster, Laird, & Rubin**

Nested Models

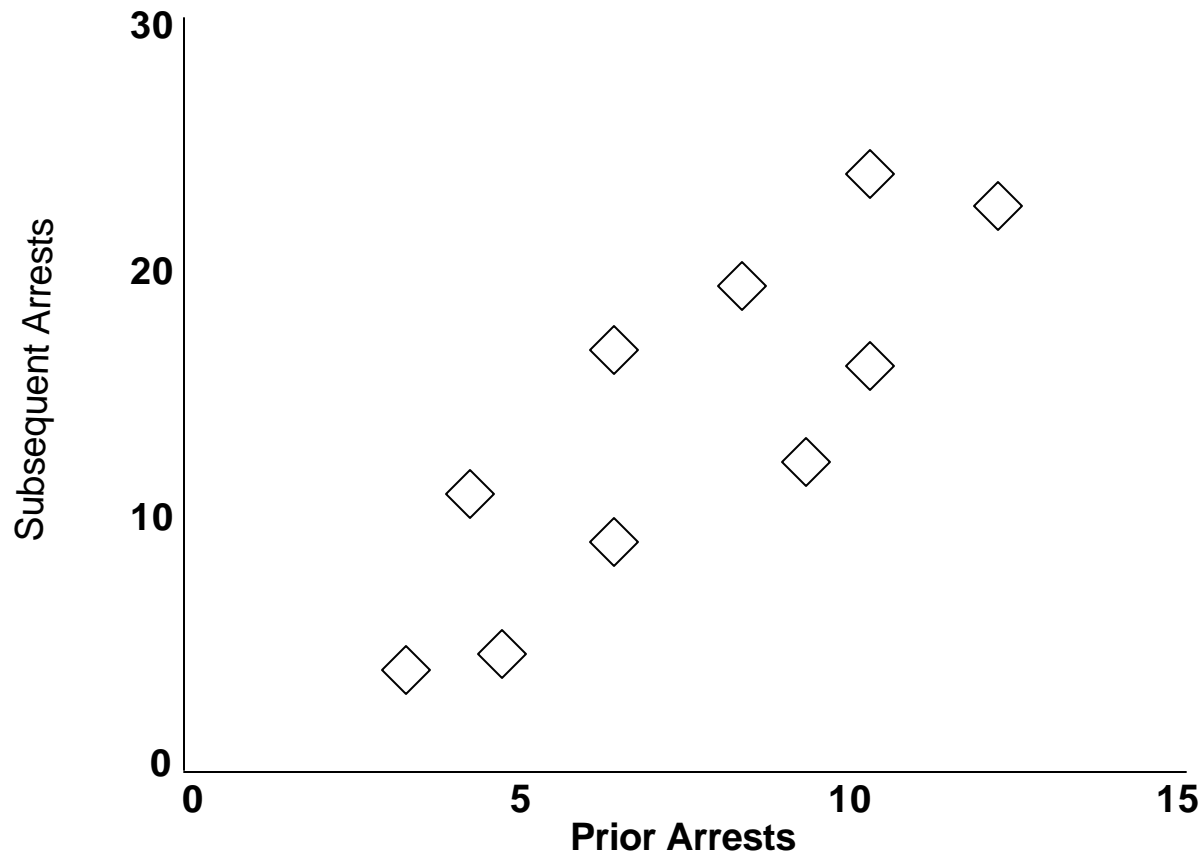
Public Policy research commonly employs nested structures. Examples:

- **People are often nested within some organizational unit (e.g., classrooms, schools, school districts, states, etc.)**
- **Repeated observations are collected on a set of individuals and the measurement occasions are not identical for all persons. Here, occasions are nested within individuals.**

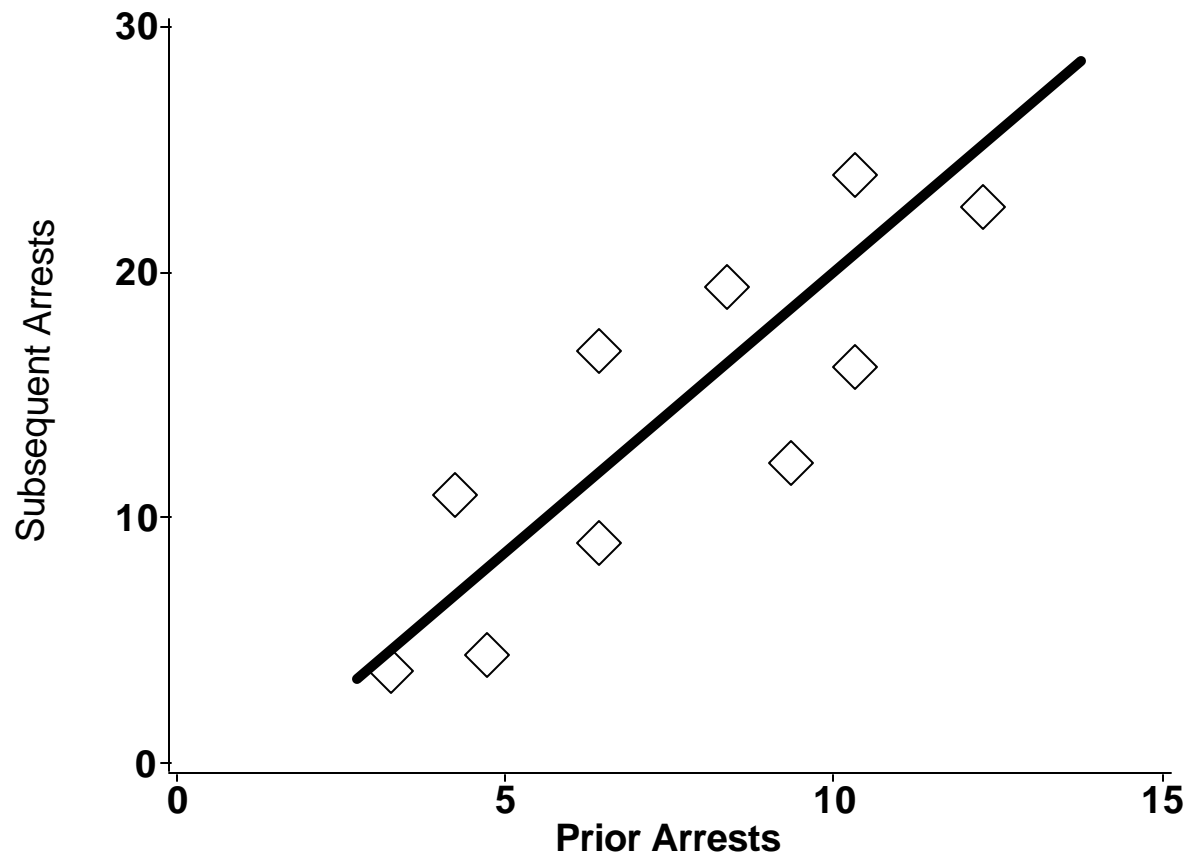
What Does HLM Do?

- **Fits a regression equation at the lowest level**
- **Lets parameters of the regression equation vary by group membership**
- **Uses higher-level variables to “explain” variation in the lower-level parameters**
- **Permits you to test for main effects and interactions, both within and between levels.**

One Probation Officer

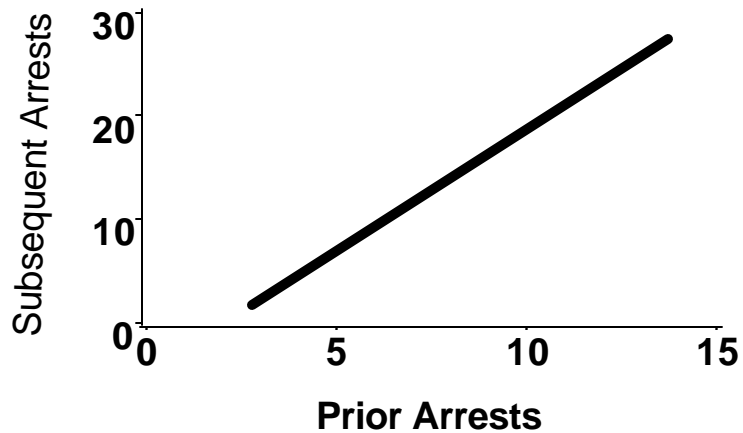


The Regression Line

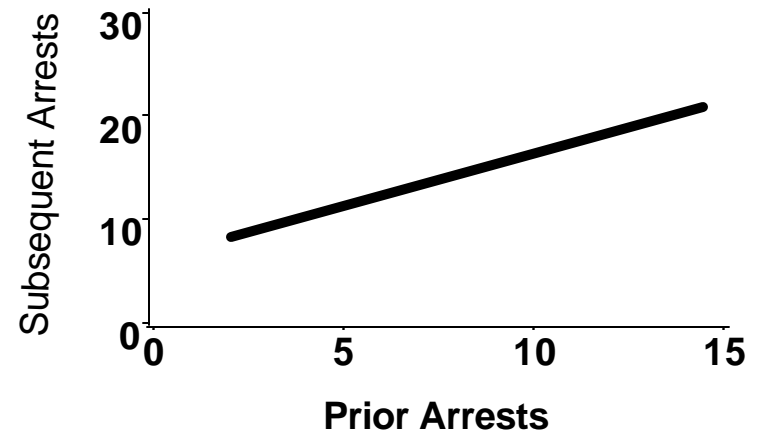


Multiple Probation Officers

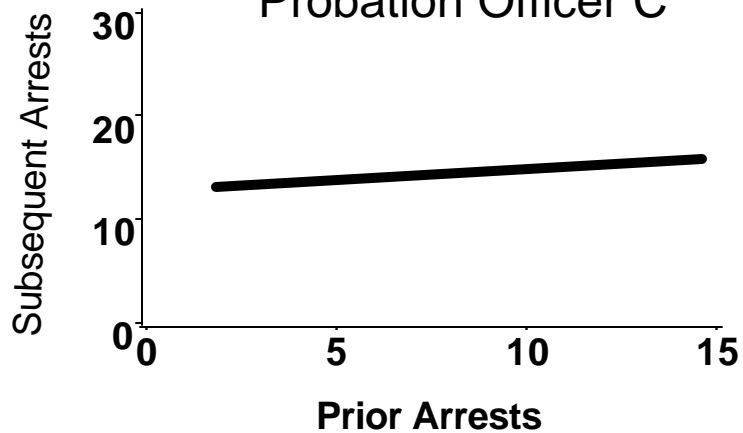
Probation Officer A



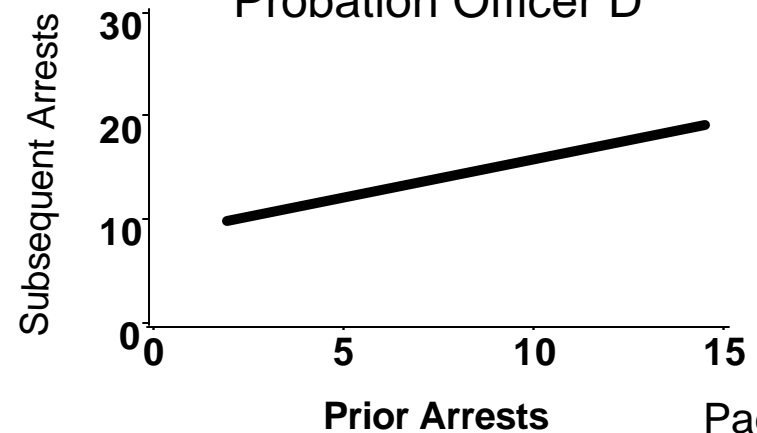
Probation Officer B



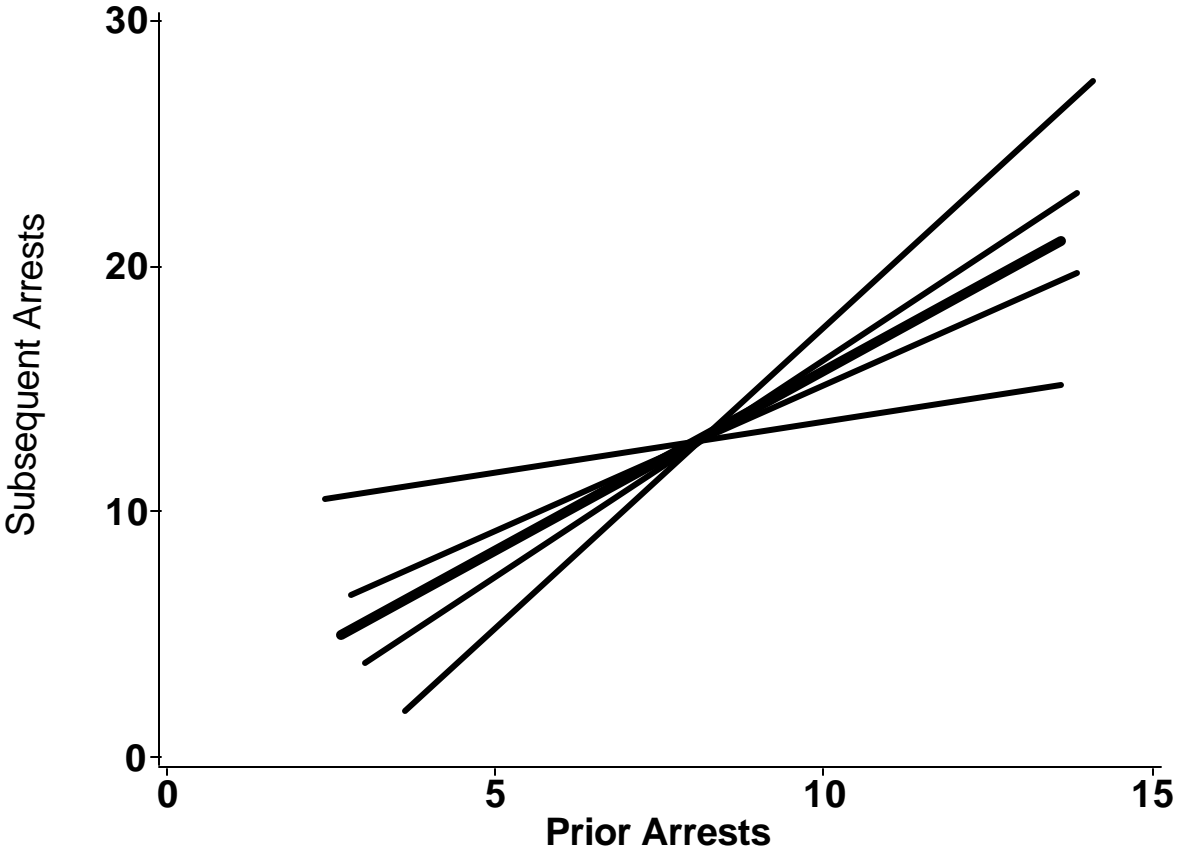
Probation Officer C



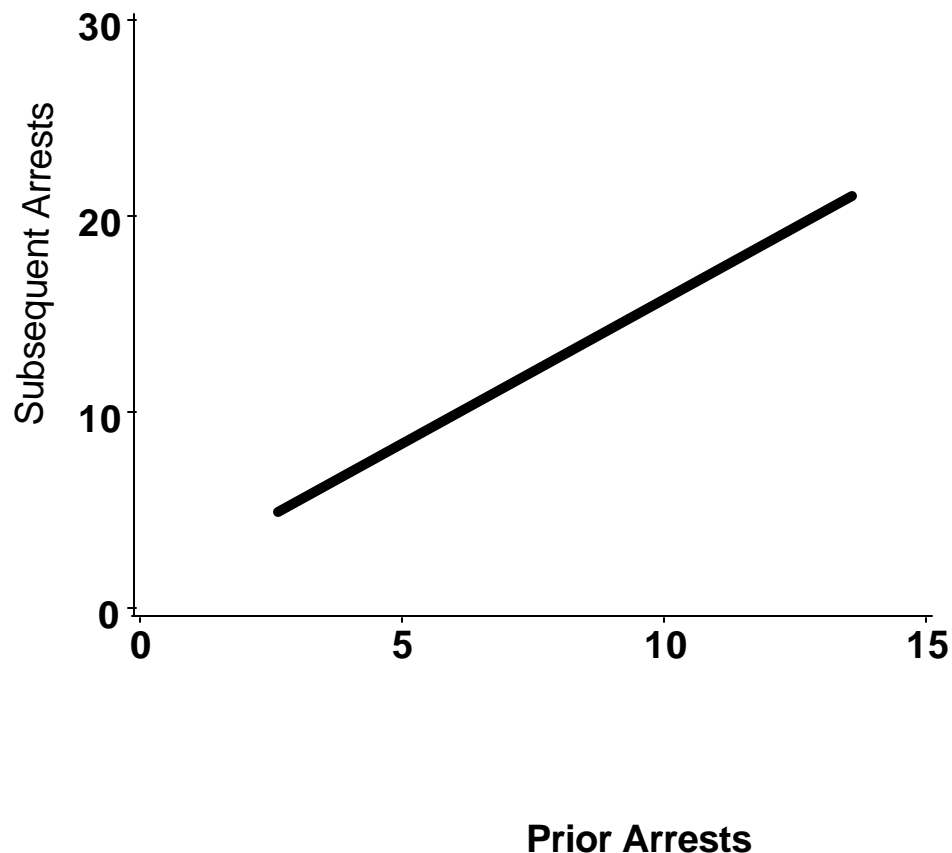
Probation Officer D



The “Average” Probation Officer



Average Probation Officer's Slope: Statistically Significant?



The challenge: We need to determine whether, based on these data, we can be reasonably certain that the slope of the “average” probation officer’s line *in the population* is anything other than horizontal.

This might be an easy task except for one crucial fact: the standard error of the slope for the “average” probation officer is in part affected by the error around the individual probation officers’ lines.

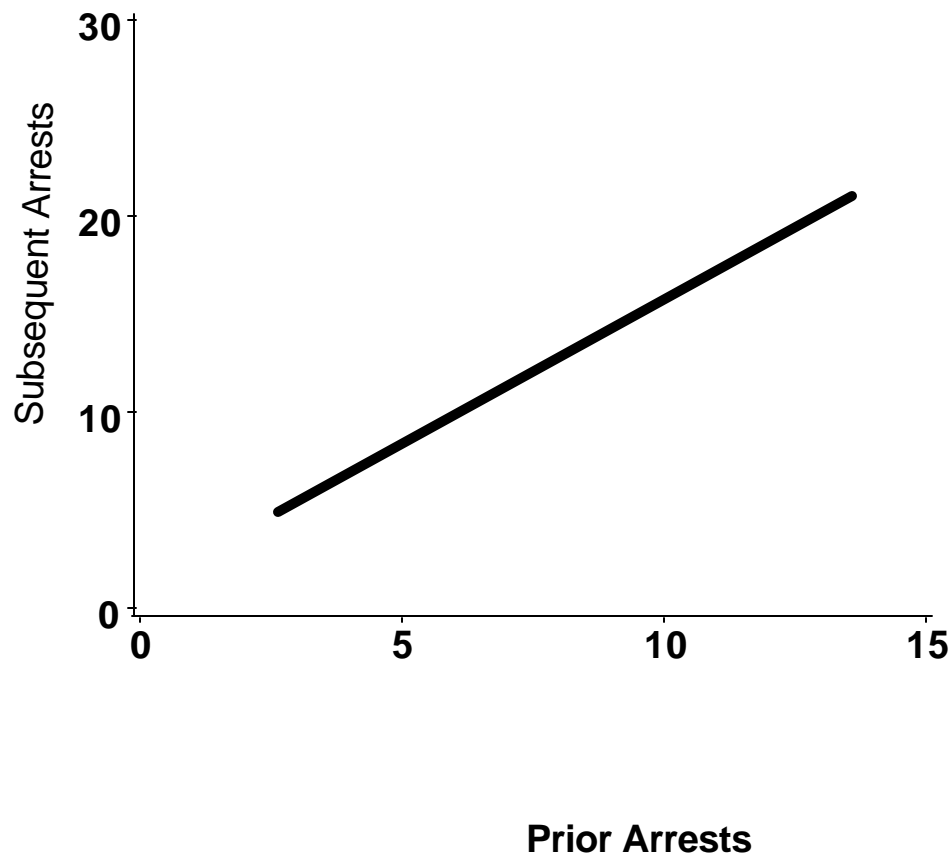
Estimation of the statistical significance of the slope is also complicated by the fact that probation may officers carry different loads (i.e., that the “sample size” is not the same across probation officers).

Prototypical HLM Table (Excerpt)

<u>Fixed Effect</u>	<u>Coefficient</u>	<u>se</u>	<u>t-ratio</u>	<u>p-value</u>
Slope	0.45	0.09	4.92	.0006

<u>Random Effects</u>	<u>Variance</u>	<u>Z-value</u>	<u>p-value</u>
Slopes	0.09	2.23	.0259
Residual	0.32		

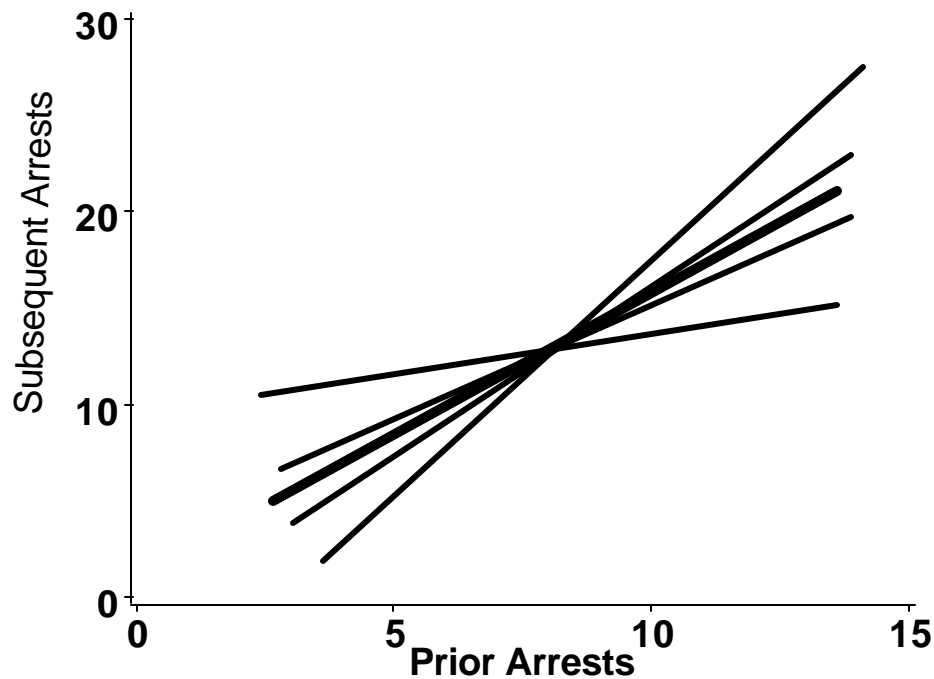
Interpreting the Table: Fixed Effects



The average probation officer in the population is estimated to have a slope of 0.45

It would be unlikely to see a slope this steep or steeper (in either direction) if the average probation officer in the population was characterized by a best-fitting regression line that was horizontal

Interpreting the Table: Random Effects



The best estimate of the variance in slopes among probation officers is 0.09

It is unlikely that the variance is zero in the population

Inputting Data for PROC MIXED

```
DATA MF1;  
  INFILE ...;  
  INPUT PROBOFF ... PARRESTS ... SARRESTS ... EXPER ...;
```

HLM in SAS

(Excluding Experience)

```
proc mixed data=mf1 covtest;
class proboff;
model sarrests = parrests;
random intercept parrests /type=un subject=proboff;
estimate 'Intercept' intercept 1;
estimate 'Slope' parrests 1;
```

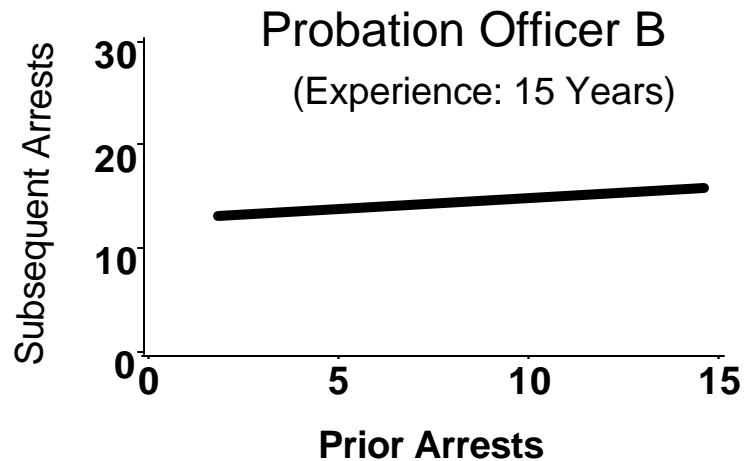
Covariance Parameter Estimates (REML)

Cov Parm	Subject	Estimate	Std Error	Z	Pr > Z
UN(1,1)	PROBOFF	11.25332475	5.08516839	2.21	0.0269
UN(2,1)	PROBOFF	-1.00313568	0.45200590	-2.22	0.0265
UN(2,2)	PROBOFF	0.08982456	0.04031276	2.23	0.0259
Residual		0.32362300	0.02304946	14.04	0.0001

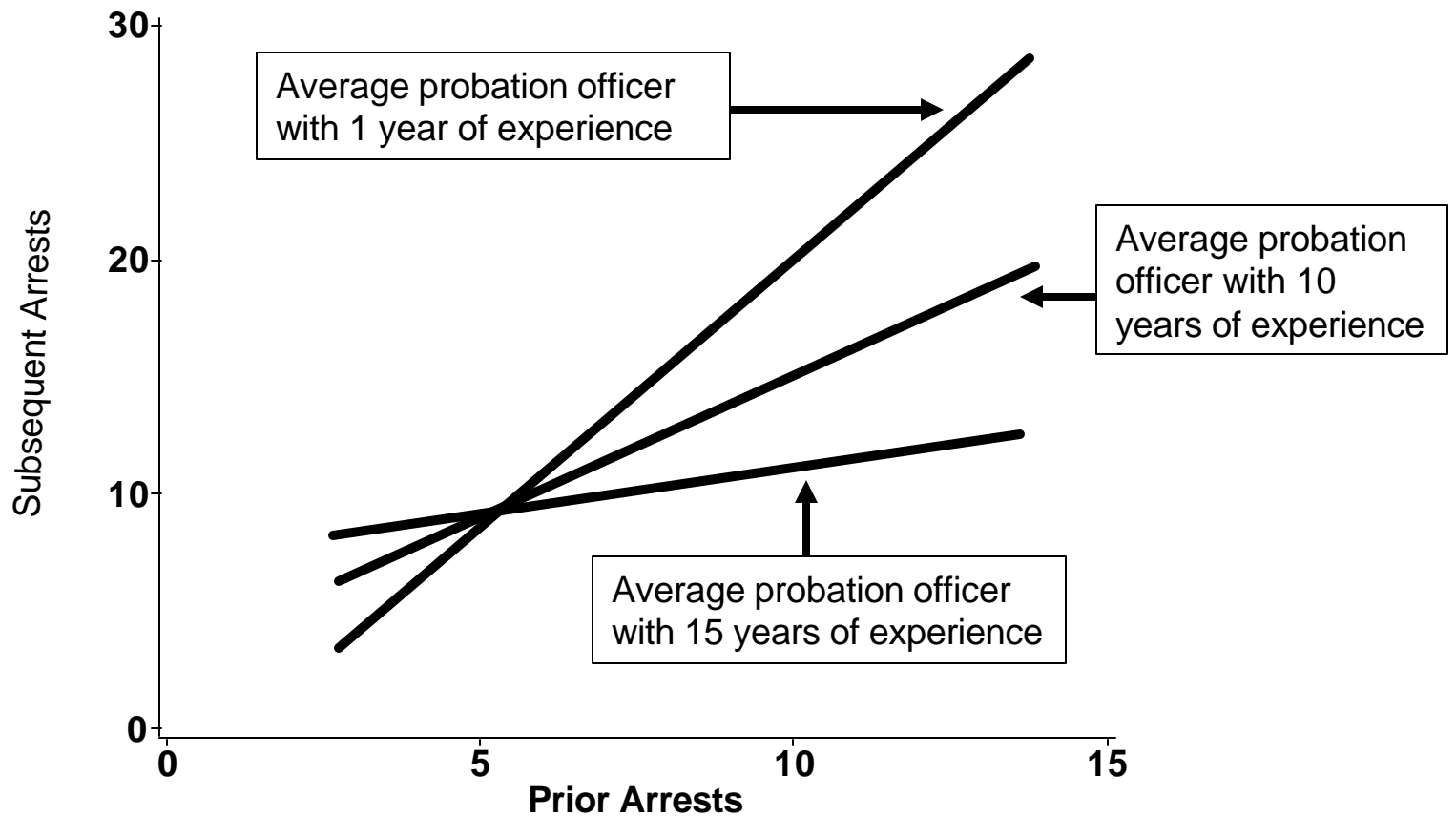
ESTIMATE Statement Results

Parameter	Estimate	Std Error	DF	t	Pr > t
Intercept	11.45666267	1.01666711	10	11.27	0.0001
Slope	0.44503620	0.09052152	10	4.92	0.0006

Factoring in Experience: Multiple Probation Officers



Factoring in Experience



HLM in SAS

(Including Experience)

```
proc mixed data=mf1 covtest;  
class proboff;  
model sarrests = parrests exper parrests*exper;  
random intercept parrests /type=un subject=proboff;  
estimate "Arr.*Exper. Int." parrests*exper 1;
```

Covariance Parameter Estimates (REML)					
Cov Parm	Subject	Estimate	Std Error	Z	Pr > Z
UN(1,1)	PROBOFF	0.00665199	0.05513797	0.12	0.9040
UN(2,1)	PROBOFF	0.00250136	0.00219751	1.14	0.2550
UN(2,2)	PROBOFF	0.00000000	.	.	.
Residual		0.31822168	0.02241200	14.20	0.0001

ESTIMATE Statement Results					
Parameter	Estimate	Std Error	DF	t	Pr > t
Arr.*Exper. Int.	-0.06010750	0.00104452	394	-57.55	0.0001

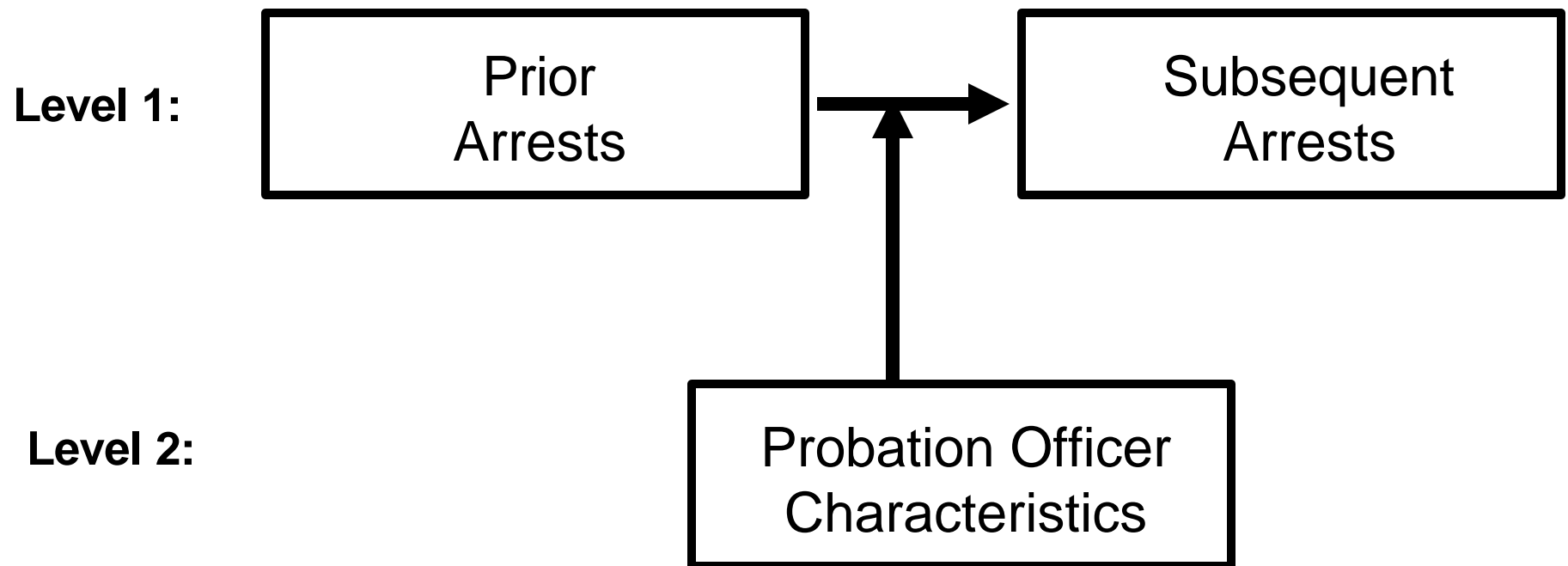
HLM Table

(Including Experience)

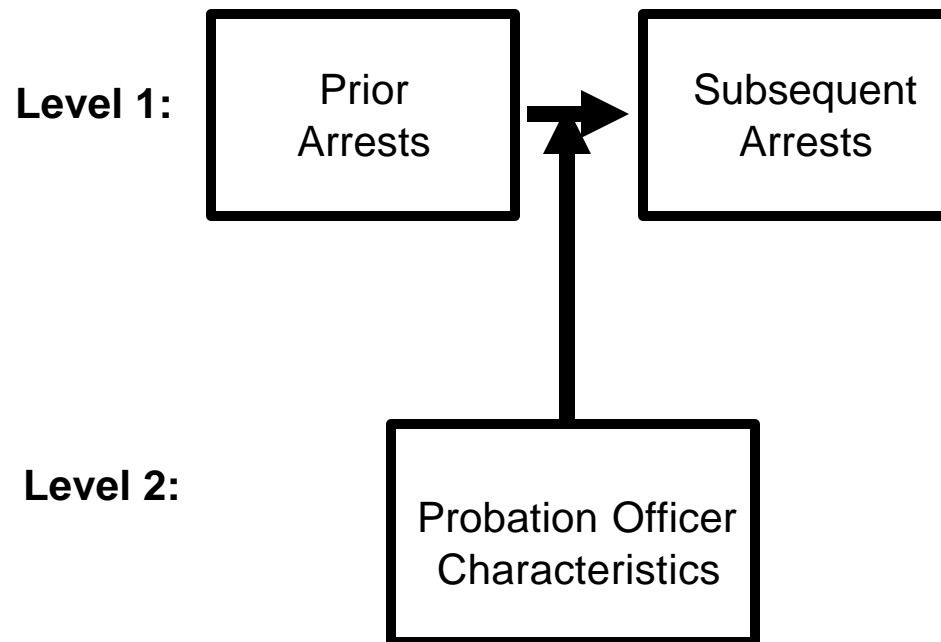
<u>Fixed Effect</u>	<u>Coefficient</u>	<u>se</u>	<u>t-ratio</u>	<u>p-value</u>
Slope*Experience	-0.06	0.001	-57.55	.0001

<u>Random Effects</u>	<u>Variance</u>	<u>Z-value</u>	<u>p-value</u>
Slopes	0.00		
Residual	0.32		

Multi-Level Causal Diagram



The “Bottom Line”



These data provide convincing evidence that clients of less experienced probation officers tend to reflect “stronger” relationships between prior and subsequent arrests

Hierarchical Linear Modeling ("HLM")

- Allows you to specify variables from *multiple* (nested) levels of analysis
- Effects can be fixed or random
- Balanced design is *not* required
- Algorithm correctly estimates the associated standard error for determination of statistical significance

Hierarchical Linear Modeling: Other Example Applications

- **Effect on treatment outcome of matching therapist characteristics to client characteristics**
- **Site effects on the relationship between client baseline characteristics and treatment outcomes in a multi-site clinical trial**
- **Analysis of individual change (“latent growth”) over time**

The Analysis of Change: Latent Growth Modeling

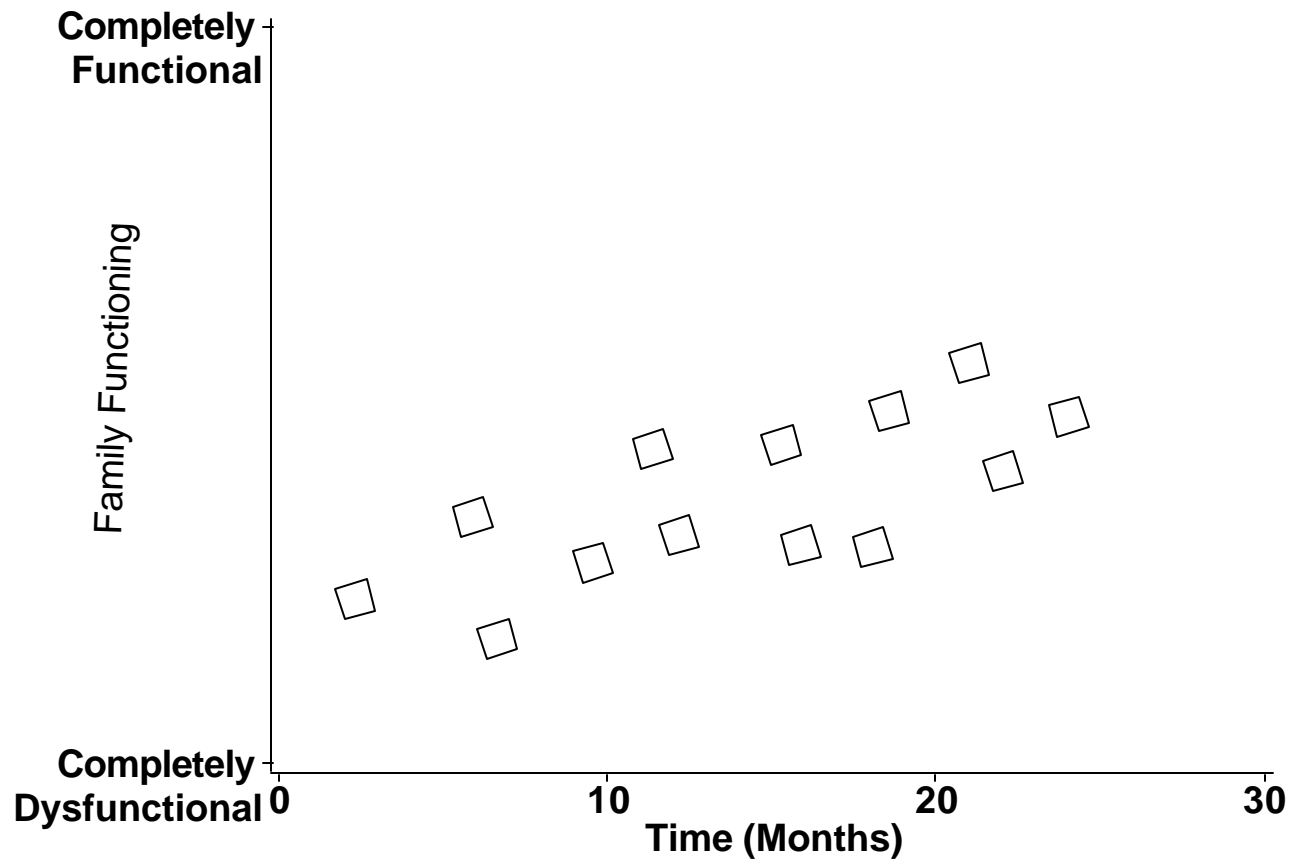
- **Still a multiple-levels problem**
- **The level-2 unit of analysis is now the organism (e.g., a specific person or a specific family)**
- **The level-1 unit of analysis is now the individual measurement point (e.g., the follow-up point)**
- **Each organism should be followed up for a minimum of three time points**

The Analysis of Change: Example¹

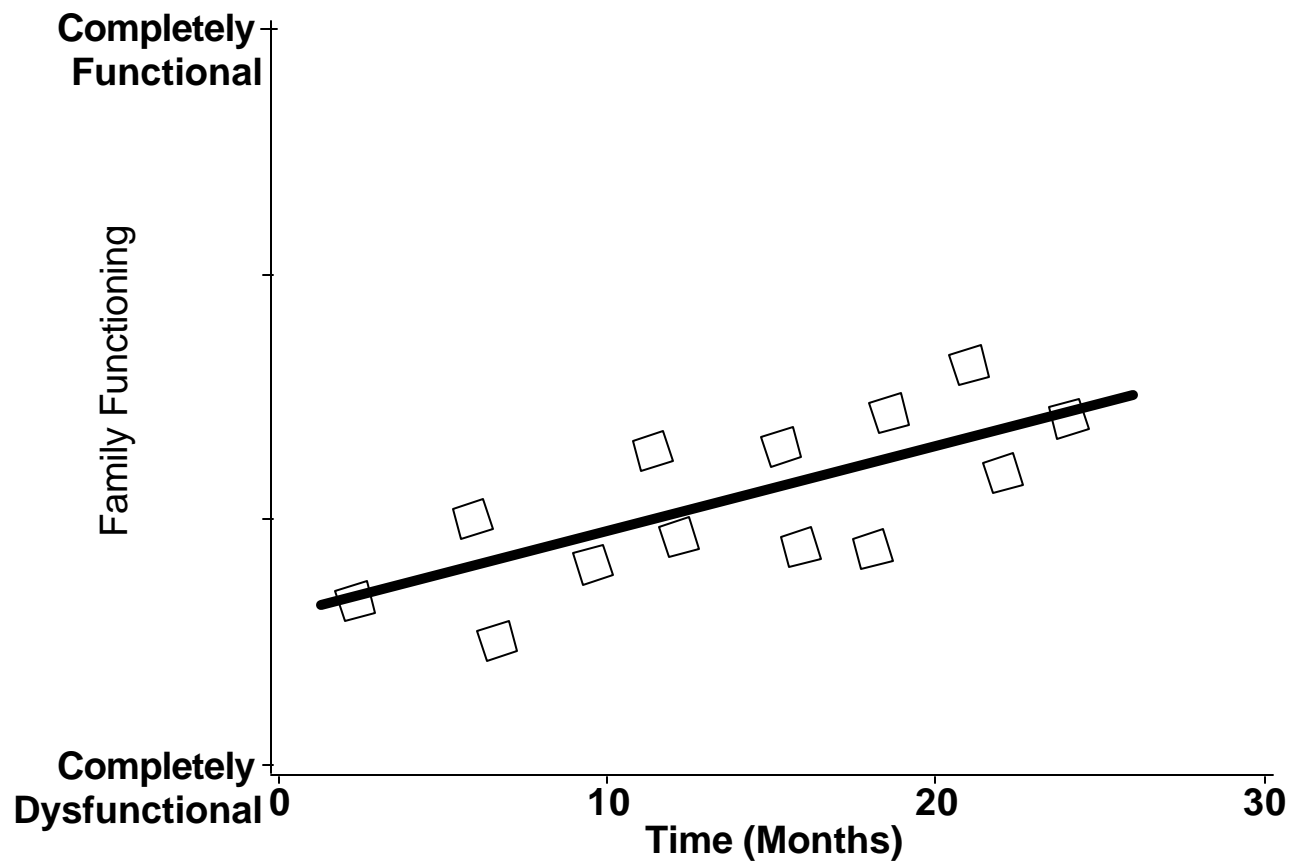
- **172 volunteer families at-risk of maladaptive parenting, child abuse, or neglect**
- **Assessed on “family functioning” over an extended period of time while in treatment**
- **Also assessed on a number of “baseline” indicators, including reported prior violence and maltreatment problems**

¹ Adapted from Willett, J.B., Ayoub, C.C., & Robinson, D. (1991). Using growth modeling to examine systematic differences in growth: An example of change in the functioning of families at risk of maladaptive parenting, child abuse, or neglect. *Journal of Consulting and Clinical Psychology*, *59*, 38-47. All data presented here have been totally fabricated for pedagogical purposes.

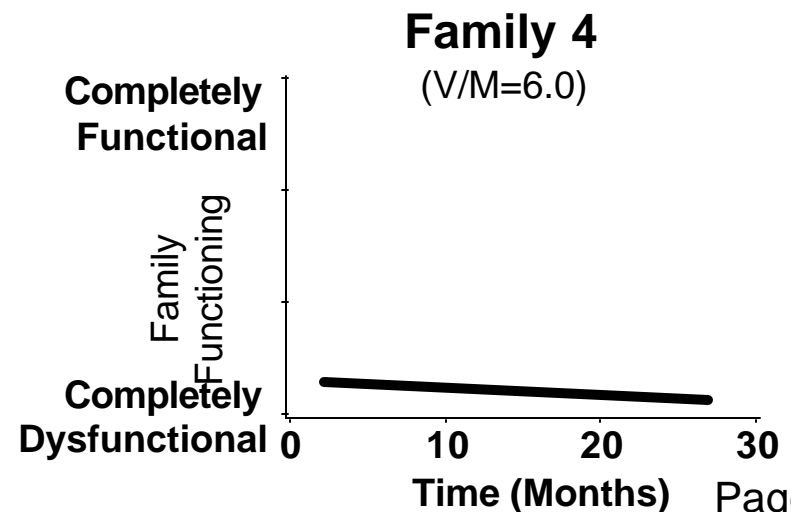
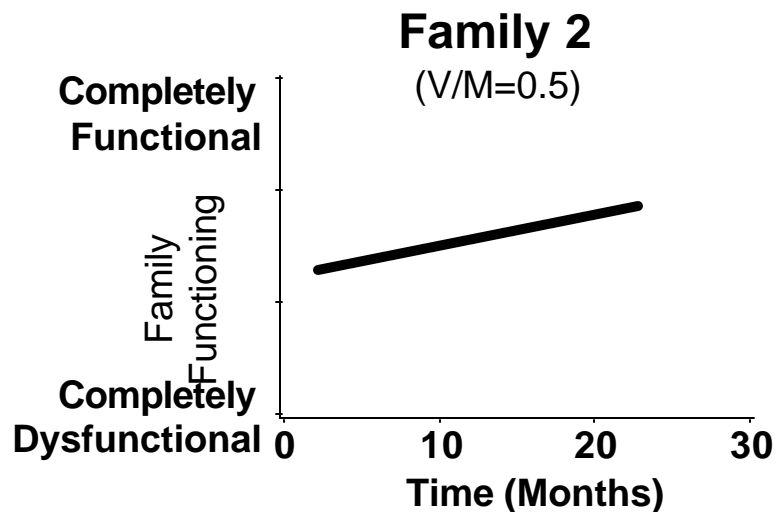
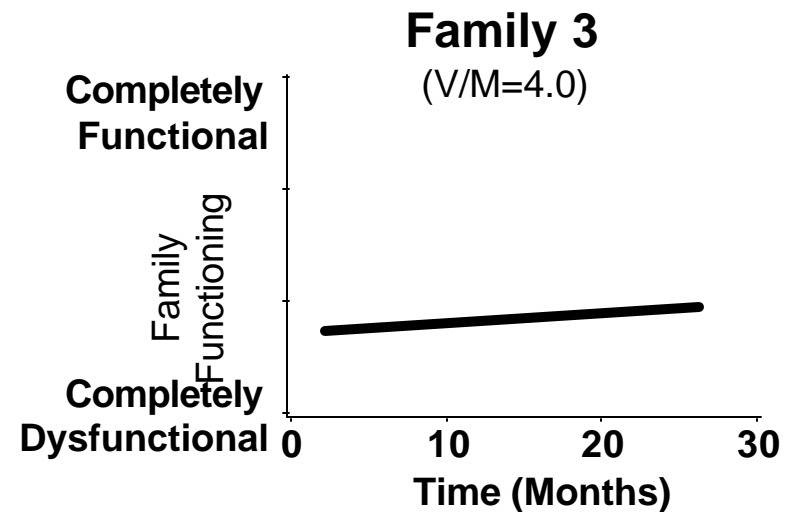
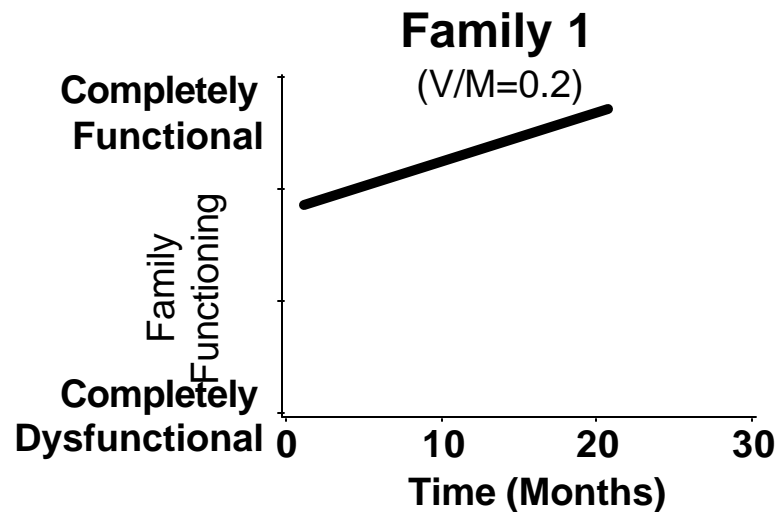
One Family



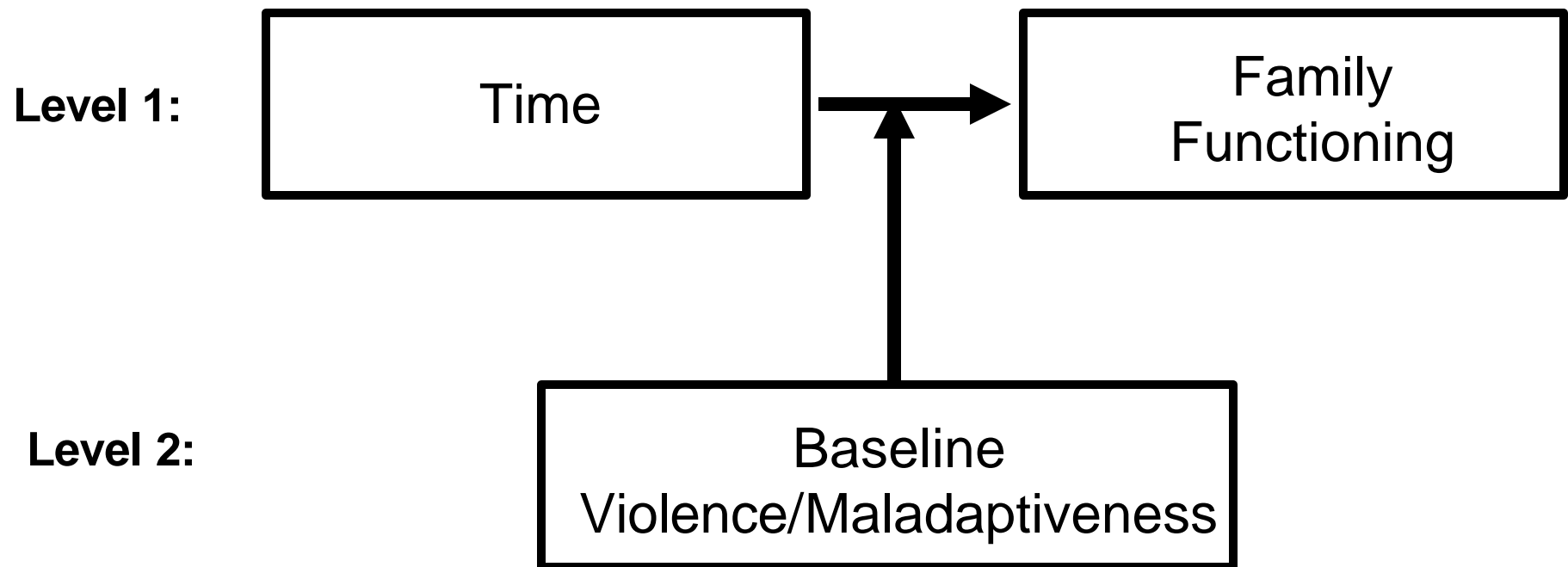
The Regression Line of One Family



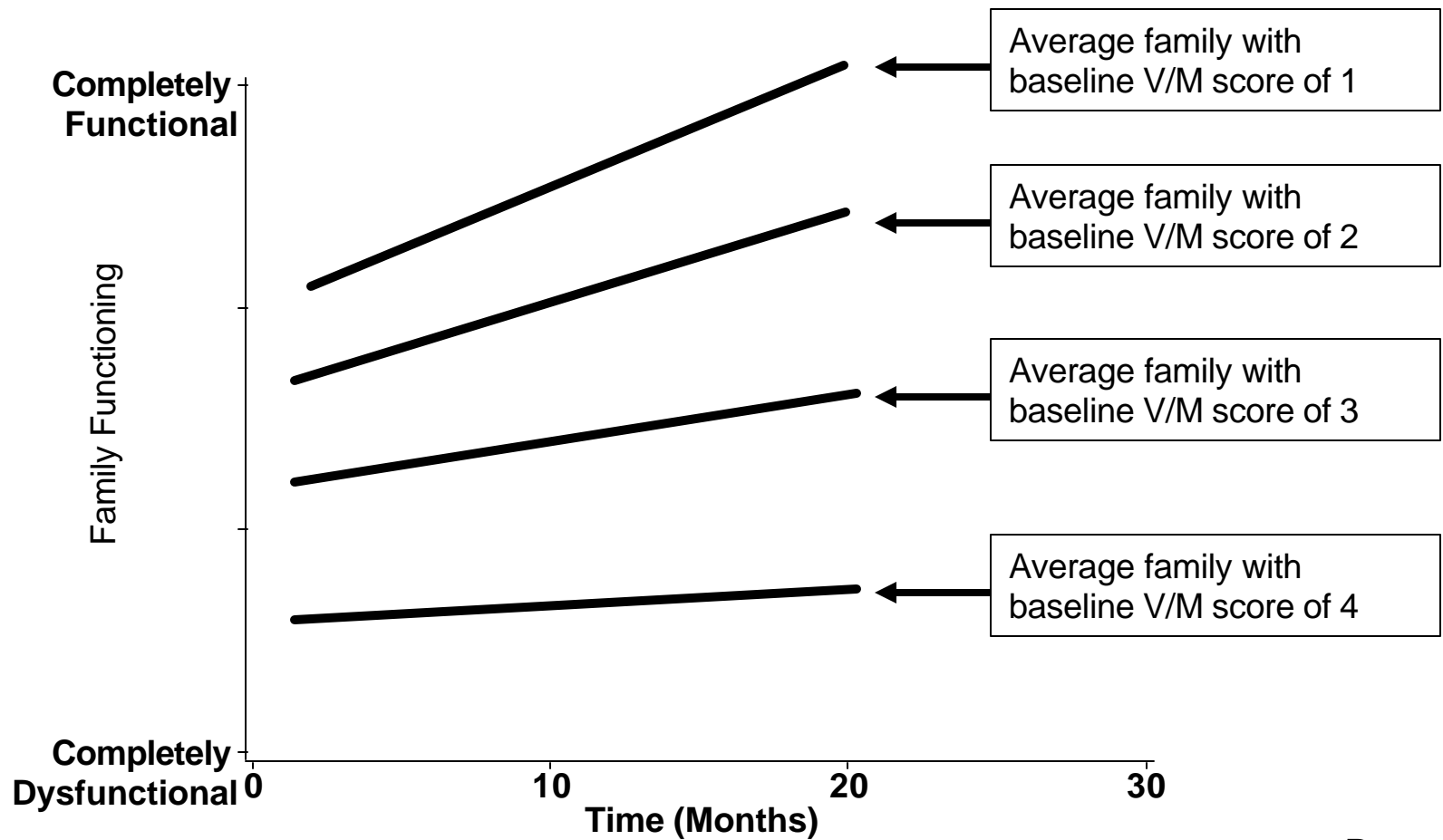
The Regression Lines of Several Families



Multi-Level Causal Diagram



Regression Lines for “Average” Families at Selected V/M Scores



Testing for the Impact of Level 2 Variables on Level 1 Intercept

```
proc mixed data=mf1 covtest;  
  class familyid;  
  model func = time vm time*vm;  
  random intercept time /type=un subject=proboff;  
  estimate "VM" vm 1;
```

The “Bottom Line”

These data provide convincing evidence that families with higher levels of baseline violence/maladaptiveness tend to be characterized by lower initial functioning level and slower recovery rates compared to families with lower levels of baseline violence/maladaptiveness.

Extensions of Hierarchical Linear Modeling

- **Can easily handle models which are non-linear in the data (but not models which are non-linear in the parameters)**
- **Has been extended to discrete outcome variables**
- **Has been extended to structural equation models (“M-Plus”)**
- **Has been extended to handle missing data**

Useful References

- **Bryk, A.S., & Raudenbush, S.W. (1992). Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park, CA: Sage.**
- **Raudenbush, S.W., & Bryk, A.S. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods (2nd Edition). Newbury Park, CA: Sage.**
- **(Less useful 😊)**
http://oak.itl.gwu.edu/FacultyFiles/Wirtz/HLM_GWIPP.ram