Utilizing Hierarchical Linear Modeling in Evaluation: *Concepts and Applications*

C.J. McKinney, M.A.

Antonio Olmos, Ph.D.

Kate DeRoche, M.A.

Mental Health Center of Denver

Evaluation 2007: Evaluation and Learning November 8th , 2007
Baltimore, MD

Overview

Hierarchical Linear Regression Concepts

Building a Model

Some Applications in Evaluation with Examples

- Program/Policy Impact (Spline Model)
- Probability of Incidence (Logistic Regression)

Very Brief Review of Misuses of HLM

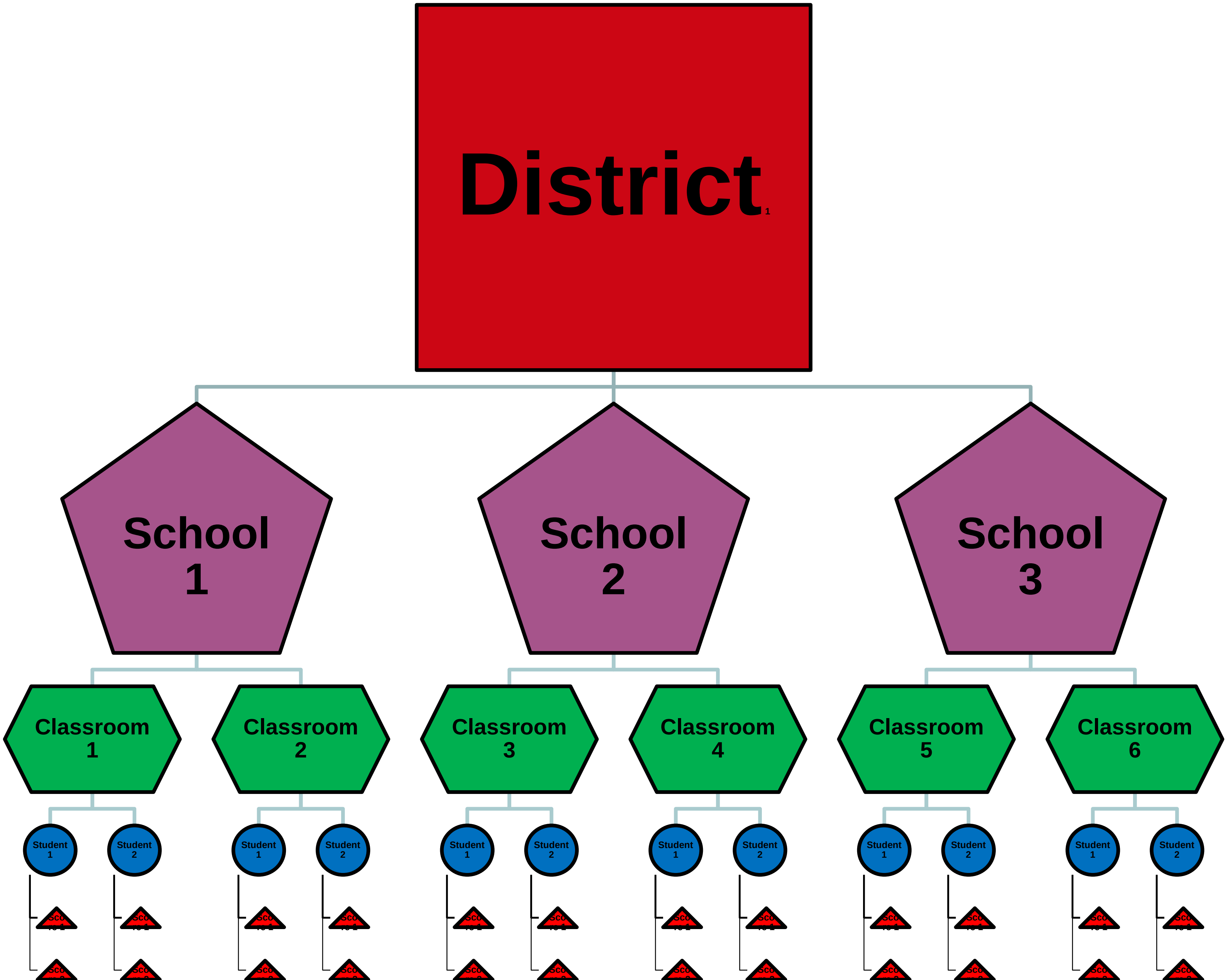
Discussion/Questions

Theoretical Rationale

Environment

Social and Behavioral researchers typically study situations where higher level factors affect lower level outcomes.

- For Example:
  - An individual worker’s error rate may be affected by the departments average workload per employee, which may be affected by the division’s product demand
  - Evaluation of student reading improvement in Colorado middle schools.



District
School 1
School 2
School 3
Classroom 1
Classroom 2
Classroom 3
Classroom 4
Classroom 5
Classroom 6
Student 1
Student 2
Student 1
Student 2
Student 1
Student 2
Student 1
Student 2
Student 1
Student 2
Student 1
Student 2

Parameter Estimation

Hierarchical linear regression models typically provide better estimates of relationships than more conventional regression models.

- Estimates:
  - Higher accuracy
  - Lower standard error

Due to taking into account hierarchical nature of the data.

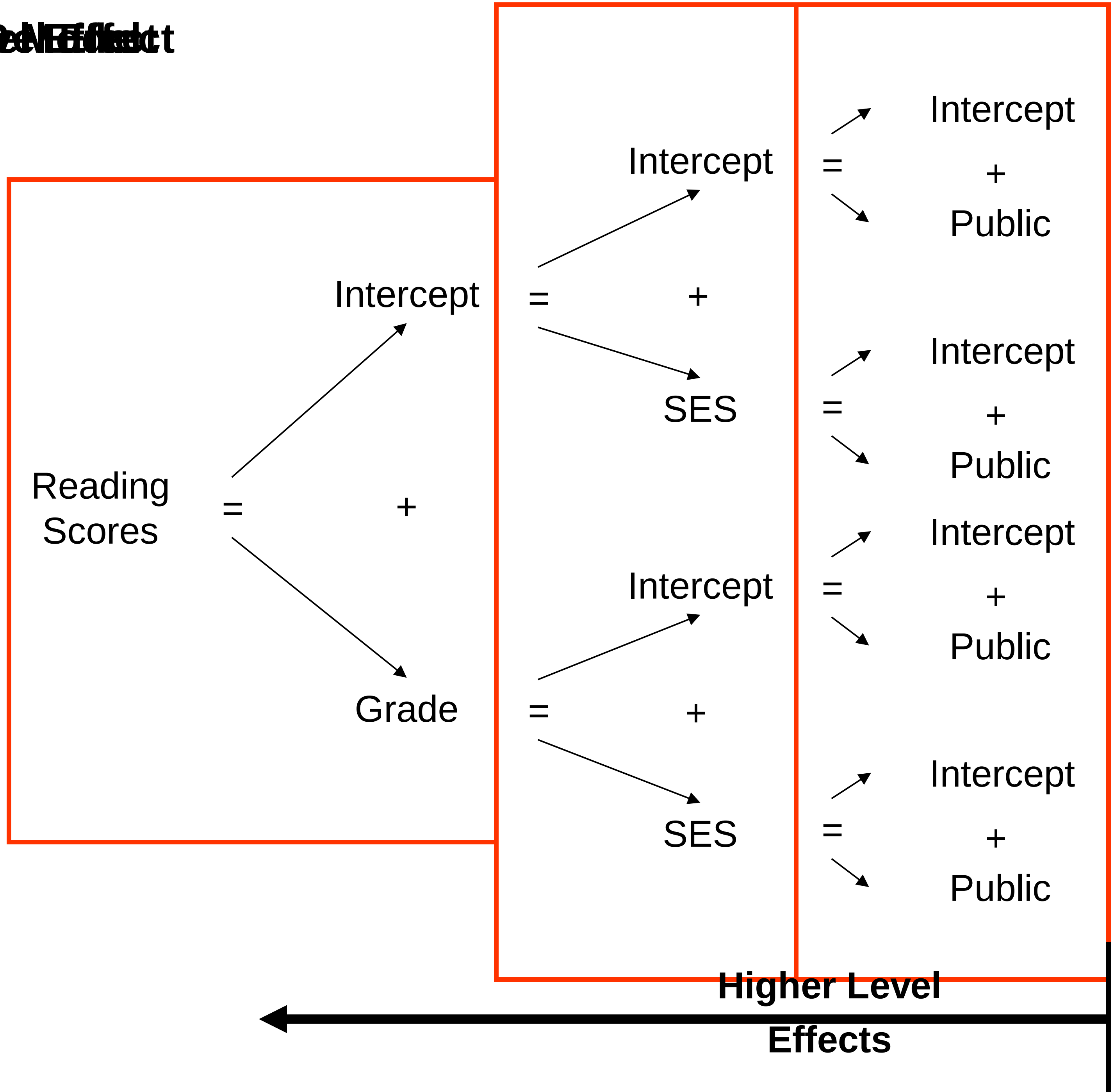
Hierarchical Linear Regression Concepts

Coefficients as Outcomes

Hierarchical linear regression models use the coefficients of regression at the lower levels as outcomes in regression at the higher levels.

- Described as “Regression on Regression”
- Older hierarchical linear modeling papers referred to the technique as random coefficient models for this reason.

[illegible]



Reading Scores
=
Intercept
+
Grade
Intercept
=
+
SES
Intercept
=
+
SES
Intercept
=
+
Public
Intercept
=
+
Public
Intercept
=
+
Public
Intercept
=
+
Public
Higher Level Effects

Estimation of Errors

Errors are estimated at each level and for each cross-level effect

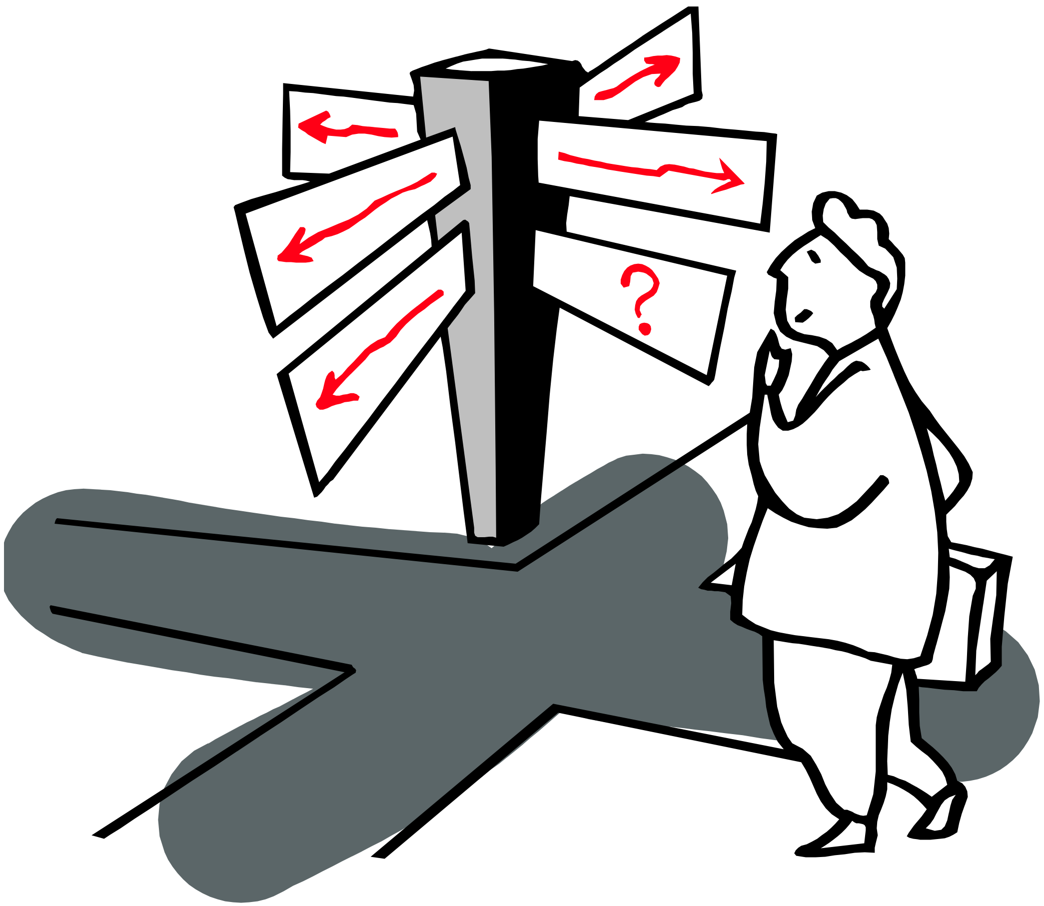
	Outcomes	Intercept	Slope	Error
LEVEL 1:	Score	$= \mathbf{Grade_{intercept}}$	$+ \mathbf{Grade_{slope}}$	$+ \varepsilon_{11}$
LEVEL 2:	$\mathrm{Grade}_{intercept}$	$= \mathbf{SES_{intercept1}}$	$+ \mathbf{SES_{slope1}}$	$+ r_{21}$
	Grade_{slope}	$= \mathbf{SES_{intercept2}}$	$+ \mathbf{SES_{slope2}}$	$+ r_{22}$
LEVEL 3:	$\mathrm{SES}_{intercept1}$	$= \mathbf{Public_{intercept1}}$	$+ \mathbf{Public_{slope1}}$	$+ u_{31}$
	SES_{slope1}	$= \mathbf{Public_{intercept2}}$	$+ \mathbf{Public_{slope2}}$	$+ u_{32}$
	$\mathrm{SES}_{intercept2}$	$= \mathbf{Public_{intercept3}}$	$+ \mathbf{Public_{slope3}}$	$+ u_{33}$
	SES_{slope2}	$= \mathbf{Public_{intercept4}}$	$+ \mathbf{Public_{slope4}}$	$+ u_{34}$

Assumptions

Hierarchical linear regression models have the same assumptions as conventional regression models

- Linear relationship
- Errors have Normal distribution
- Errors have Equal Variances
- Errors are Independent

Questions???



Building an HLM model

DECA Scores

Score 1: Measures teacher perception of child's protective factors through 3 subgroups:

- Initiative
- Self-Control
- Attachment

Score 2: Measure of teacher's concern for the child's behavior.

Scores taken at program admission (beginning of school year) and after program discharge (end of school year) N = 536

DECA Scores

Factor of Interest:

Pre/Post Scoring

Developmental Factors:

Age and Gender

Unconditional Model

LEVEL 1: $\text{Score} = \boldsymbol{\pi_0} + \boldsymbol{\varepsilon}$

LEVEL 2: $\pi_0 = \boldsymbol{\beta_{00}} + \mathbf{r_0}$

Unconditional Model

Deviance: 5502 with 2 parameters

Total Variance $= \varepsilon + r_0 = 30.2 + 42.7 = \boxed{\mathbf{72.9}}$

Variance accounted for by children: **59%**

$$\boxed{r_0/(\varepsilon + r_0) = 0.59}$$

Full Model

<u>*LEVEL 1:*</u> Score $= \boldsymbol{\pi_0} + \boldsymbol{\pi_1}$ **(Post) +ε**

<u>*LEVEL 2:*</u> $\pi_0 = \boldsymbol{\beta_{00}} + \boldsymbol{\beta_{01}}$**(Female)** + $\boldsymbol{\beta_{02}}$**(Age)** + $\mathbf{r_0}$

$\pi_1 = \boldsymbol{\beta_{10}} + \mathbf{r_1}$

<u>Deviance:</u> 5426 with 3 parameters

$\boldsymbol{\chi^2_0}$ = 5502 – 5426 = 76 => with df = 1, **p <0.005**

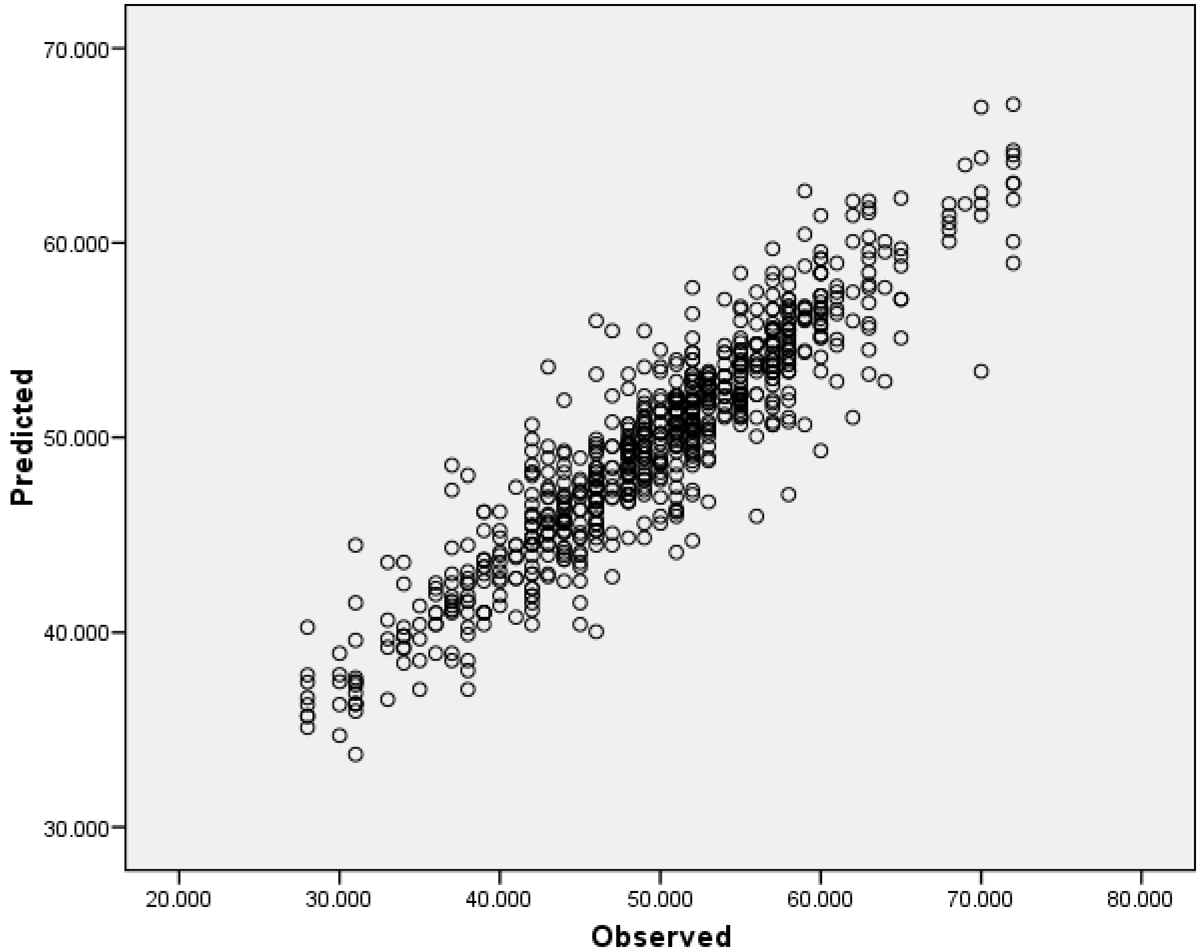
Full better than Unconditional

<u>Reduction of Error:</u> $= 1 - (\varepsilon + r_0)_{full}/(\varepsilon + r_0) = 1 - 56.7/72.9$

= 1 – 0.78 = 0.22 = **22%**

Overall Model Fit

- $R^2 = 0.85$



Parameter Estimates

POST: At the end of the program the level of protective factors increases on average by **2.6**** points over the level at the beginning of the program.

AGE: Consistent with developmental theory the level of protective factors increases by **1.4**** points per year of age.

Gender: Consistent with developmental theory the level of protective factors is greater for females than males by **3.8**** points on average.

** - Significantly different from 0 with $p < 0.001$

Questions???

Hierarchical Linear Modeling Examples

MHCD Recovery Marker Inventory

Recovery Marker Inventory

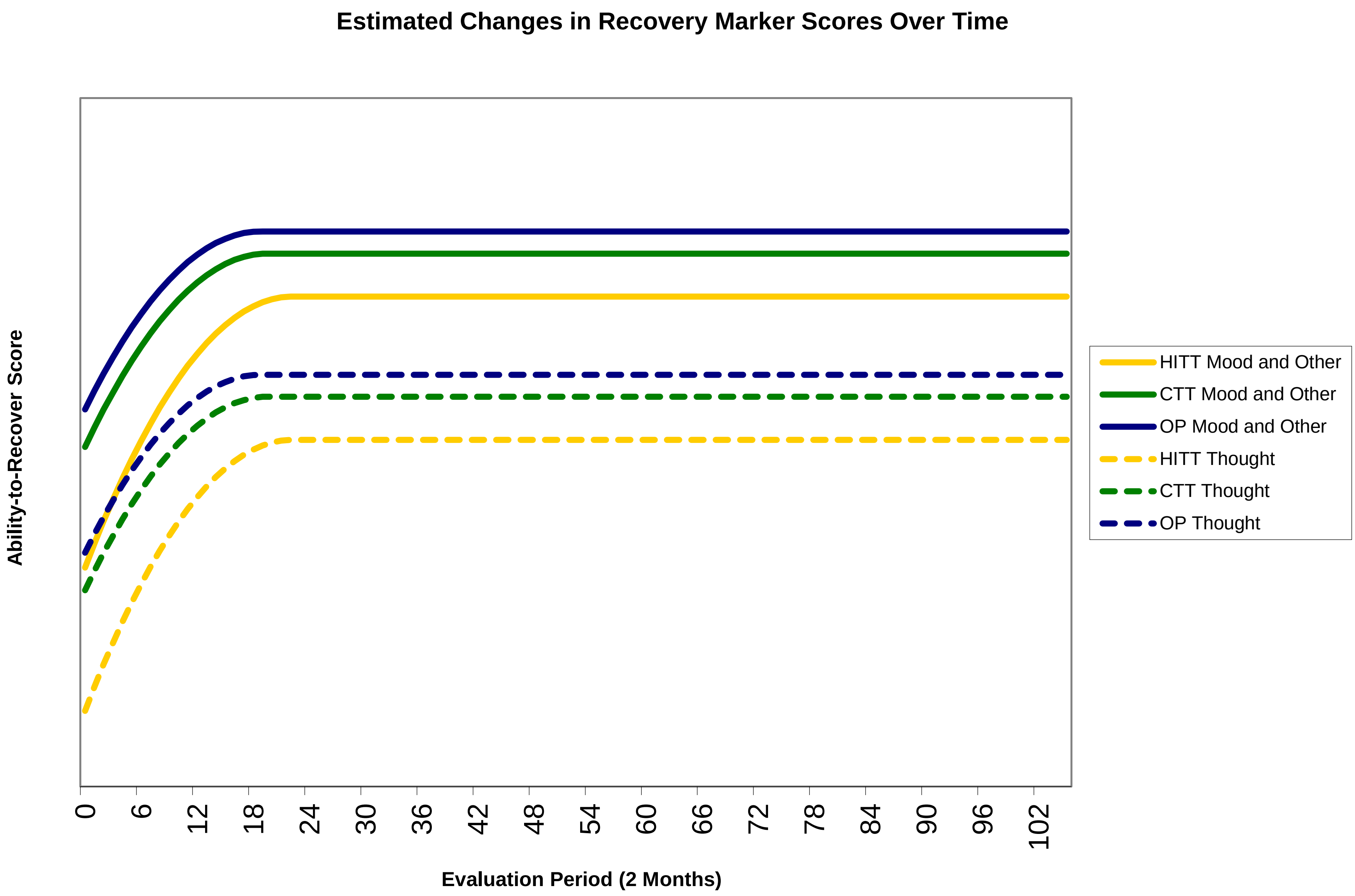
- Consist of a scaled score from six indicators of mental health recovery.
  - Employment
  - Learning/Training
  - Housing
  - Active Growth / Orientation
  - Symptom Interference
  - Service Participation

Recovery Marker Inventory

- Scores were converted from raw form to an ability score utilizing an Item Response Theory (IRT) Partial Credit model for ability estimation.
  - An increase in the ability score, indicates an increase the overall factors that support mental health recovery

Recovery Marker Inventory

- Markers are collected every 2 months on each consumer by the case managers and clinicians.



Estimated Changes in Recovery Marker Scores Over Time
Ability-to-Recover Score
0
6
12
18
24
30
36
42
48
54
60
66
72
78
84
90
96
102
Evaluation Period (2 Months)
HITT Mood and Other
CTT Mood and Other
OP Mood and Other
HITT Thought
CTT Thought
OP Thought

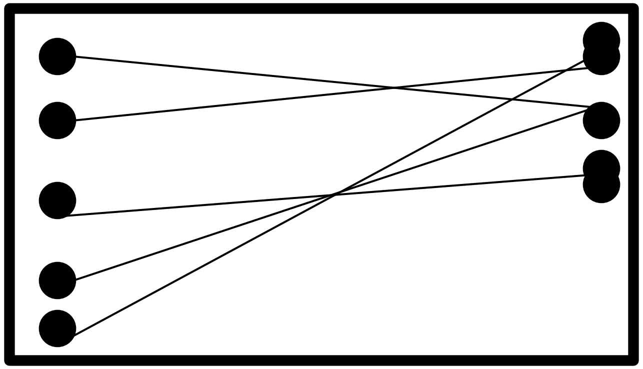
HLM Evaluation Applications

Program/Policy Impact

Typical Questions:

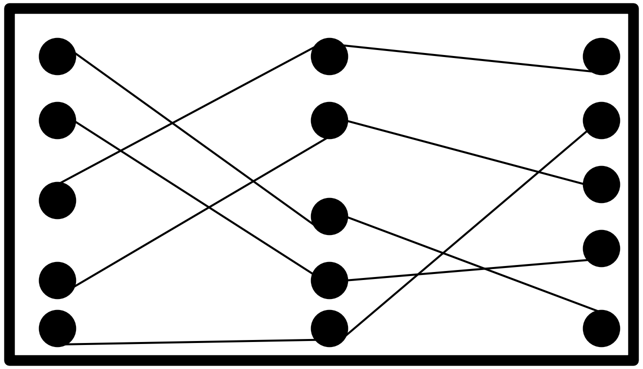
- **Does the program affect performance?**
- **Which programs are more effective?**
- **What factors affect performance?**

Pre-Post



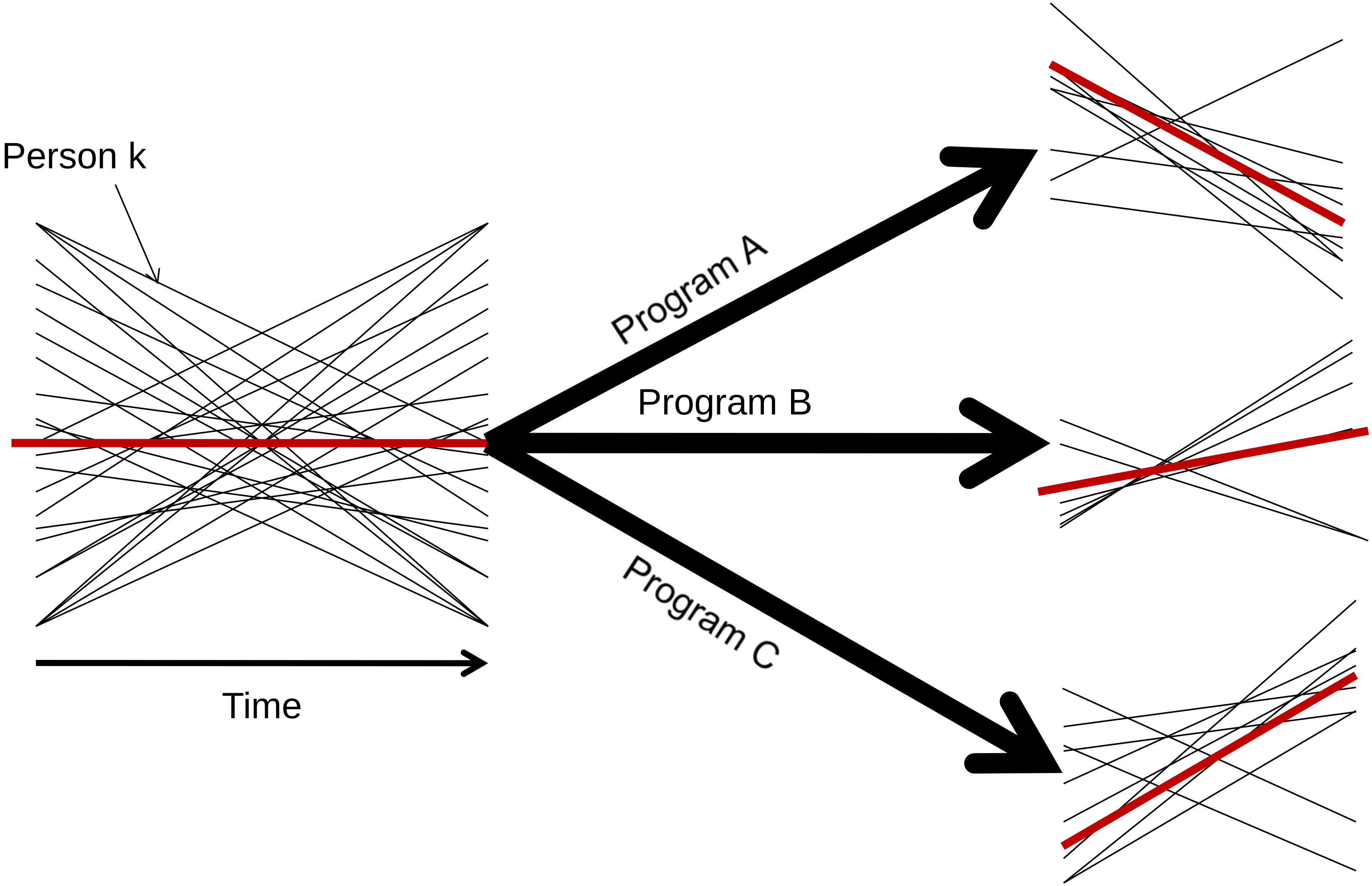
Time

> 2 Time points



Time

Program/Policy Impact



Program/Policy Impact

- The philosophy of MHCD is that all individuals with mental illness can and many do recover.

- This resulted in a policy within our adult treatment teams that as individuals recover they can be moved to a lower level team, which opens spaces for more individuals at the higher levels of treatment.

- A question arose as to whether these changes in the level of treatment affected their recovery level?

TO TEAM

FROM TEAM

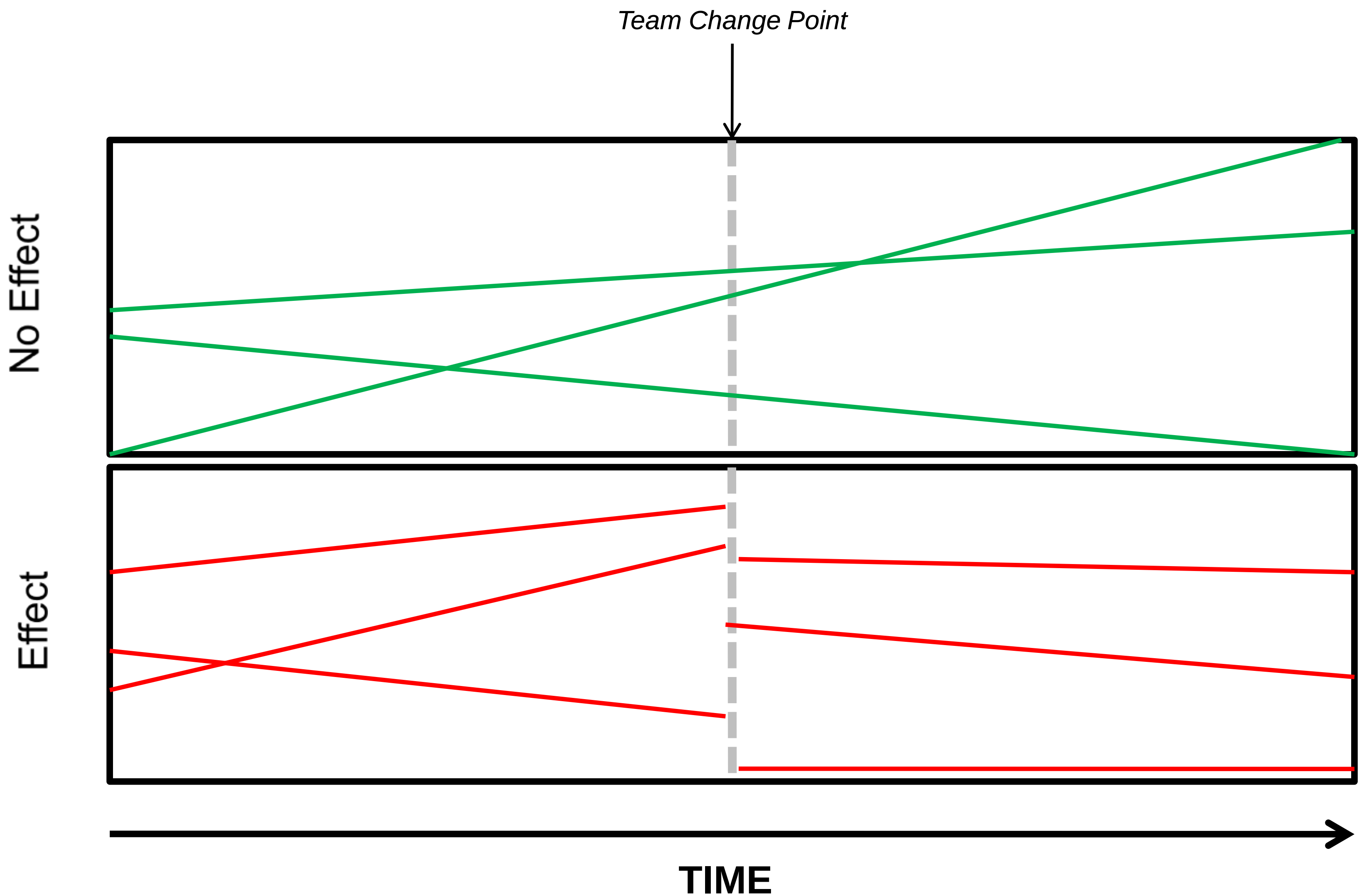
FROM TEAM \ TO TEAM	EHITT	HITT 1	HITT 2	HIIATT	HITT 3	ILT	SLT	CTT 1	CTT 2	CTT 3	OPT	OBRA	EXEC
EHITT	0	0	0	0	0	0	0	0	0	0	0	0	0
HITT 1	0	0	8	2	9	5	7	17	35	17	0	1	10
HITT 2	0	2	0	1	6	2	13	2	86	8	5	6	5
HIIATT	0	4	1	0	3	3	68	12	6	44	3	5	6
HITT 3	0	15	12	1	0	5	3	8	8	11	1	0	12
ILT	0	5	4	4	7	0	13	119	13	12	4	0	4
SLT	0	7	3	22	3	5	0	17	11	69	4	3	9
CTT 1	0	2	3	1	0	30	3	0	11	5	43	0	1
CTT 2	0	15	54	3	4	2	7	6	0	10	72	2	5
CTT 3	0	8	9	9	3	9	26	4	7	0	41	3	0
OPT		0	0	0	0	0	2	3	3	8	0	0	0
OBRA	0	0	0	0	0	0	0	0	0	1	0	0	0
EXEC	0	1	0	0	0	0	0	0	0	0	0	0	0

Increased Treatment Intensity	N=264	*21.43%*
Same Treatment Intensity	N=107	*8.69%*
Decreased Treatment Intensity	N=787	*63.88%*
Transferred to OBRA	N=20	*1.62%*
Inactive Consumer	N=52	*4.22%*

Total N = 1232

Average time in previous team 490 Days

The Model



The Model

LEVEL 1: RMIScore $= \boldsymbol{\pi_0 + \pi_1 (Time) + \pi_2 (Time^2) + \pi_3 (CHGInt) + \pi_4 (PostTime) + \pi_5 (PostTime^2) + \varepsilon}$

LEVEL 2:

$\pi_0 = \boldsymbol{\beta_{00} + \beta_{01}(MOOD) + \beta_{02}(THOUGHT) + r_0}$

$\pi_1 = \boldsymbol{\beta_{10} + \beta_{11}(MOOD) + \beta_{12}(THOUGHT) + r_1}$

$\pi_2 = \boldsymbol{\beta_{20} + \beta_{21}(MOOD) + \beta_{22}(THOUGHT) + r_2}$

$\pi_3 = \boldsymbol{\beta_{30} + \beta_{31}(MOOD) + \beta_{32}(THOUGHT) + r_3}$

$\pi_4 = \boldsymbol{\beta_{40} + \beta_{41}(MOOD) + \beta_{42}(THOUGHT) + r_4}$

$\pi_5 = \boldsymbol{\beta_{50} + \beta_{51}(MOOD) + \beta_{52}(THOUGHT) + r_5}$

Where:

Time – The time period the outcomes were obtained in number of months since admission.

CHGInt – The direction and magnitude of the team change (indicated adjustment to slope if service change occurred).

PostTime – Same as Time with values only for those after a team change occurred(Indicates a difference in slope after team change).

MOOD/THOUGHT - An indicator variable related to a consumer having a mood or thought disorder.

The Results

LEVEL 1: RMIScore $= \pi_0 + \pi_1 (\text{Time}) + \pi_2 (\text{Time}^2) + \pi_3 (\text{CHGInt}) + \pi_4 (\text{PostTime}) + \pi_5 (\text{PostTime}^2) + \varepsilon$

LEVEL 2:

$\pi_0 = \beta_{00} + \beta_{01}(\text{MOOD}) + \beta_{02}(\text{THOUGHT}) + r_0$

$\pi_1 = \beta_{10} + \beta_{11}(\text{MOOD}) + \beta_{12}(\text{THOUGHT}) + r_1$

$\pi_2 = \beta_{20} + \beta_{21}(\text{MOOD}) + \beta_{22}(\text{THOUGHT}) + r_2$

$\pi_3 = \beta_{30} + \beta_{31}(\text{MOOD}) + \beta_{32}(\text{THOUGHT}) + r_3$

$\pi_4 = \beta_{40} + \beta_{41}(\text{MOOD}) + \beta_{42}(\text{THOUGHT}) + r_4$

$\pi_5 = \beta_{50} + \beta_{51}(\text{MOOD}) + \beta_{52}(\text{THOUGHT}) + r_5$

Parameter	**B00**	**B02**	**B10**	**B20**
Estimate	4.72	-0.34	0.012	-0.0002
SE	0.033	0.04	0.0029	0.00005
p-value	<0.001	<0.001	<0.001	<0.001

Parameter	**R0**	**R1**	**R2**
Estimate	1.04	0.084	0.001
χ^2(*df*)	1789(*524*)	899(*527*)	912(*527*)
p-value	<0.001	<0.001	<0.001

Conclusions

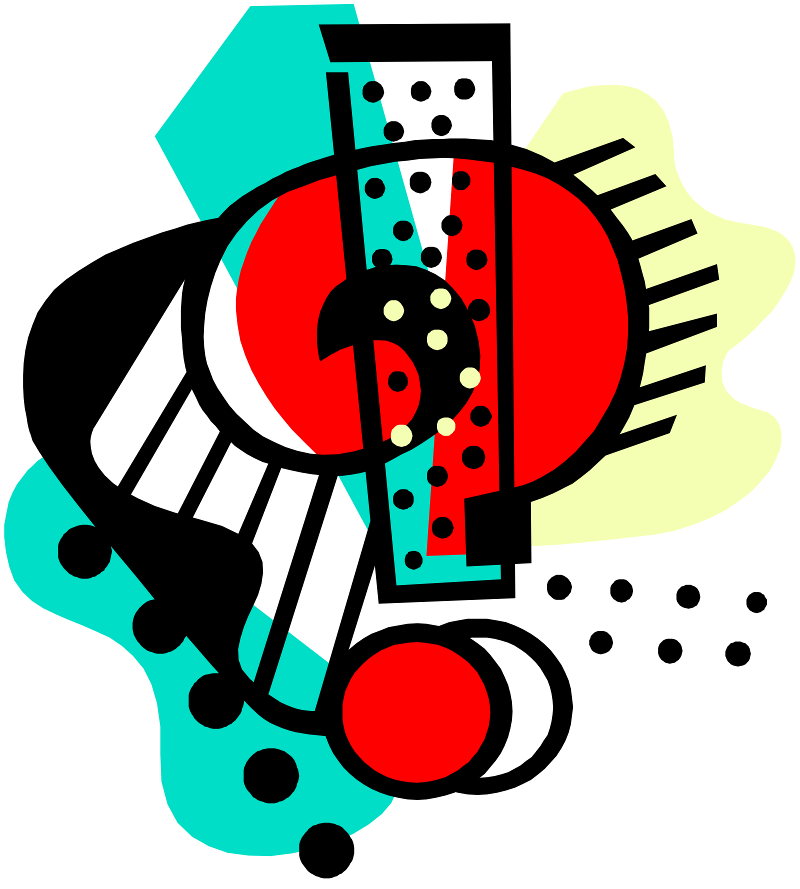
1. All of the parameters related to the team change period demonstrated no significant differences from the pre-change period.

2. The typical parameters related to time overall, and the mood and thought disorder parameters both provided results consistent with previous models.

3. This indicates that when a consumer is moved from one team to another, in either direction, they appear to keep the same level of recovery and rate of change in recovery.

4. This provides evidence to support the current practice of clinicians moving consumers to higher levels or lower levels of service as clinically indicated, as it does not affect the recovery supports for the consumer.

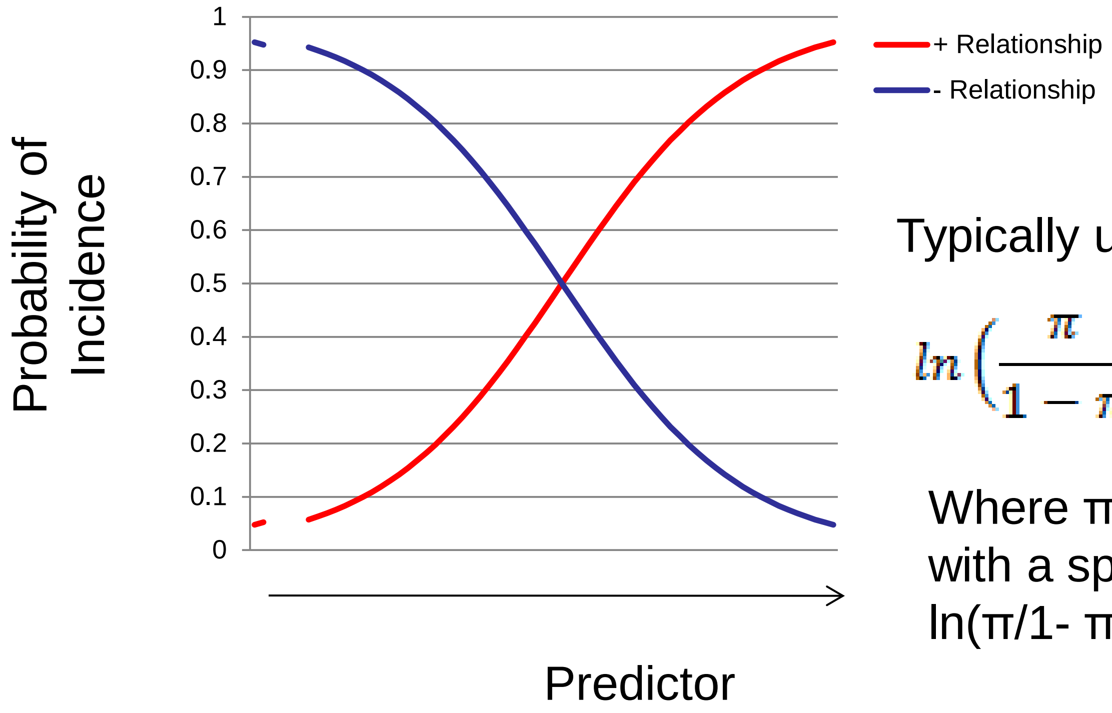
Questions???



Modeling Rate of Incidence

Typical Questions:

- Does the program affect the rate of incidence of an event?
- What factors affect the rate of incidence of an event?



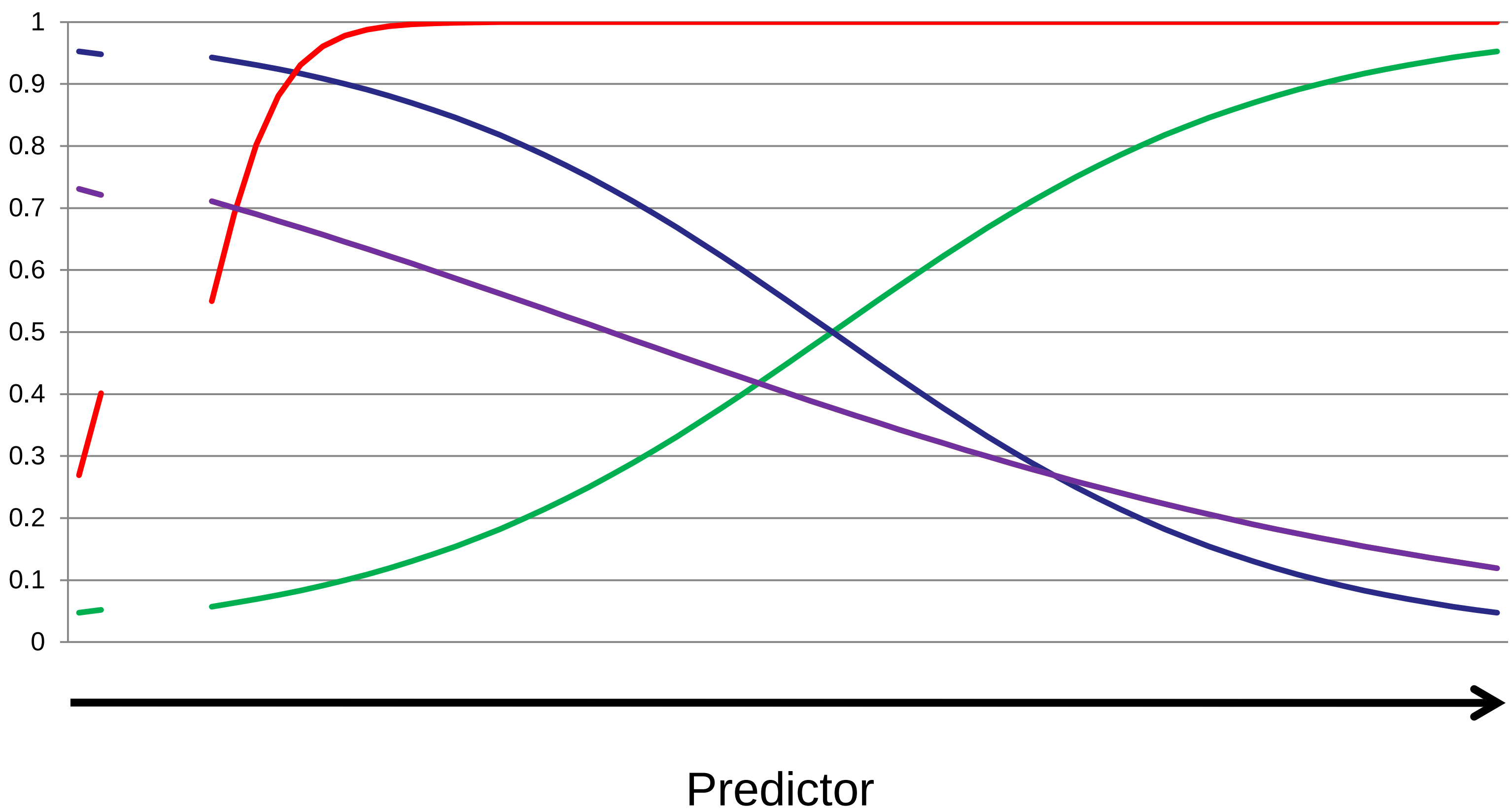
Typically use a Logistic Regression Model

$$ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1(x) + \varepsilon$$

Where π is the proportion of individuals with a specified characteristic, and ln(π/1- π) is the log-odds or logit.

Modeling Rate of Incidence

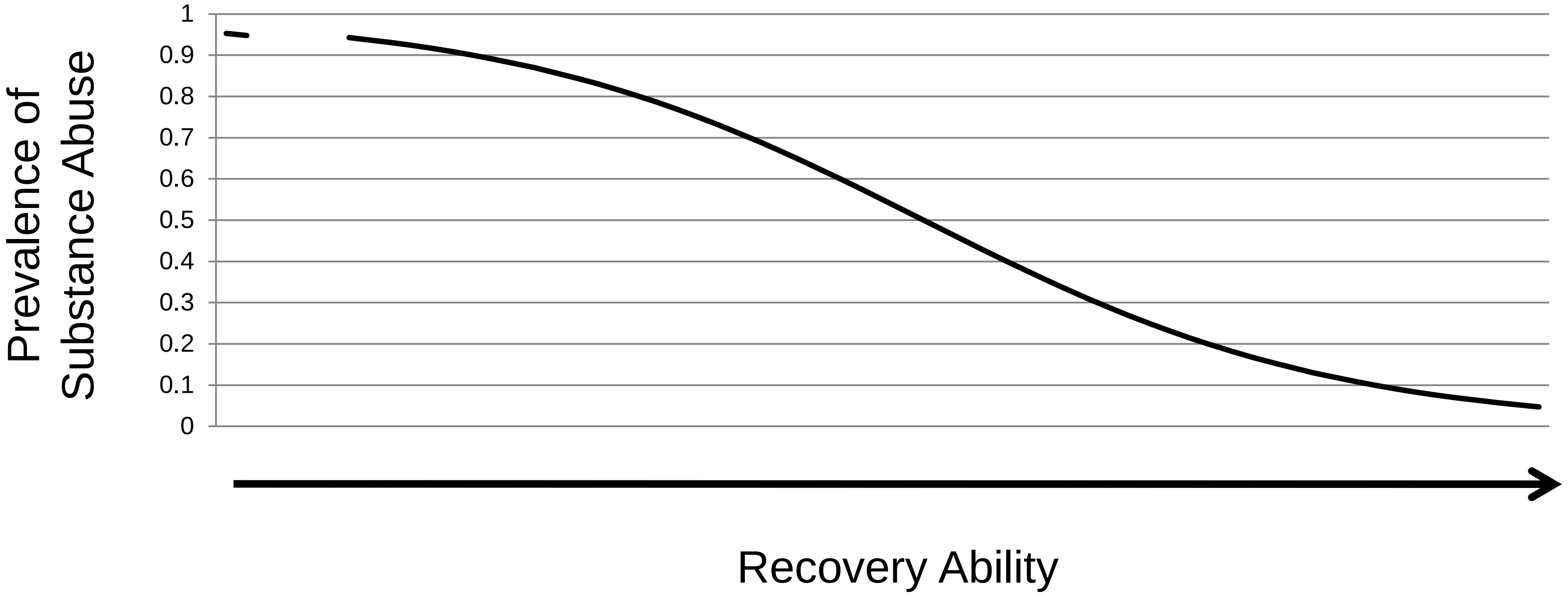
The logistic regression model can be recast into the HLM framework by simply allowing rate of incidence to vary across higher level units.



Relationship of the Prevalence of Severe Substance Abuse to Level of Recovery Ability

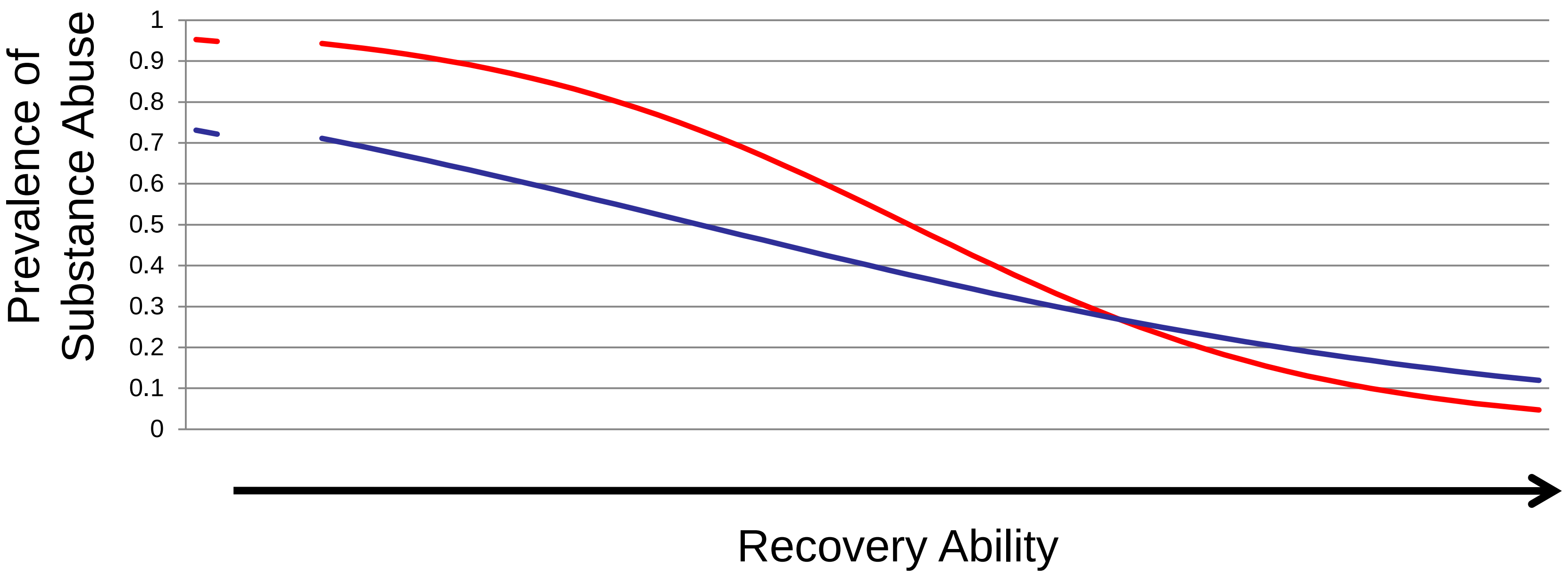
The Question: Does the prevalence of substance abuse change with the recovery ability?

- We assume that as the factors that support recovery increase, the prevalence of the abuse of substances will go down.



Relationship of the Prevalence of Severe Substance Abuse to Level of Recovery Ability

- We also assume that the rate of change varies across each individual, making the HLM model appropriate.
    - On the consumer level, we looked at whether the consumer was in a drug treatment team or not.
    - Those in a drug treatment team were expected to have a higher intercept and steeper decrease in prevalence of substance abuse.



The Model

<u>LEVEL 1:</u> $\ln(\pi/1-\pi) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}\,\textbf{(RMIScore)} + \boldsymbol{\varepsilon}$

<u>LEVEL 2:</u> $\beta_0 = \boldsymbol{\varphi_{00}} + \boldsymbol{\varphi_{01}}\textbf{(DrugTeam)} + \boldsymbol{\varphi_{02}}\textbf{(Mood)} + \boldsymbol{\varphi_{03}}\textbf{(Thought)} + \mathbf{r_0}$

$\beta_1 = \boldsymbol{\varphi_{10}} + \boldsymbol{\varphi_{11}}\textbf{(DrugTeam)} + \boldsymbol{\varphi_{12}}\textbf{(Mood)} + \boldsymbol{\varphi_{13}}\textbf{(Thought)} + \mathbf{r_1}$

Where: **ln(π/1-π)** – The log-odds of the prevalence of substance abuse.

RMIScore – The recovery marker score for each outcome for indicator of substance abuse.

DrugTeam – An indicator variable related to whether consumer was in drug treatment team or not.

MOOD/THOUGHT - An indicator variable related to a consumer having a mood or thought disorder.

Results

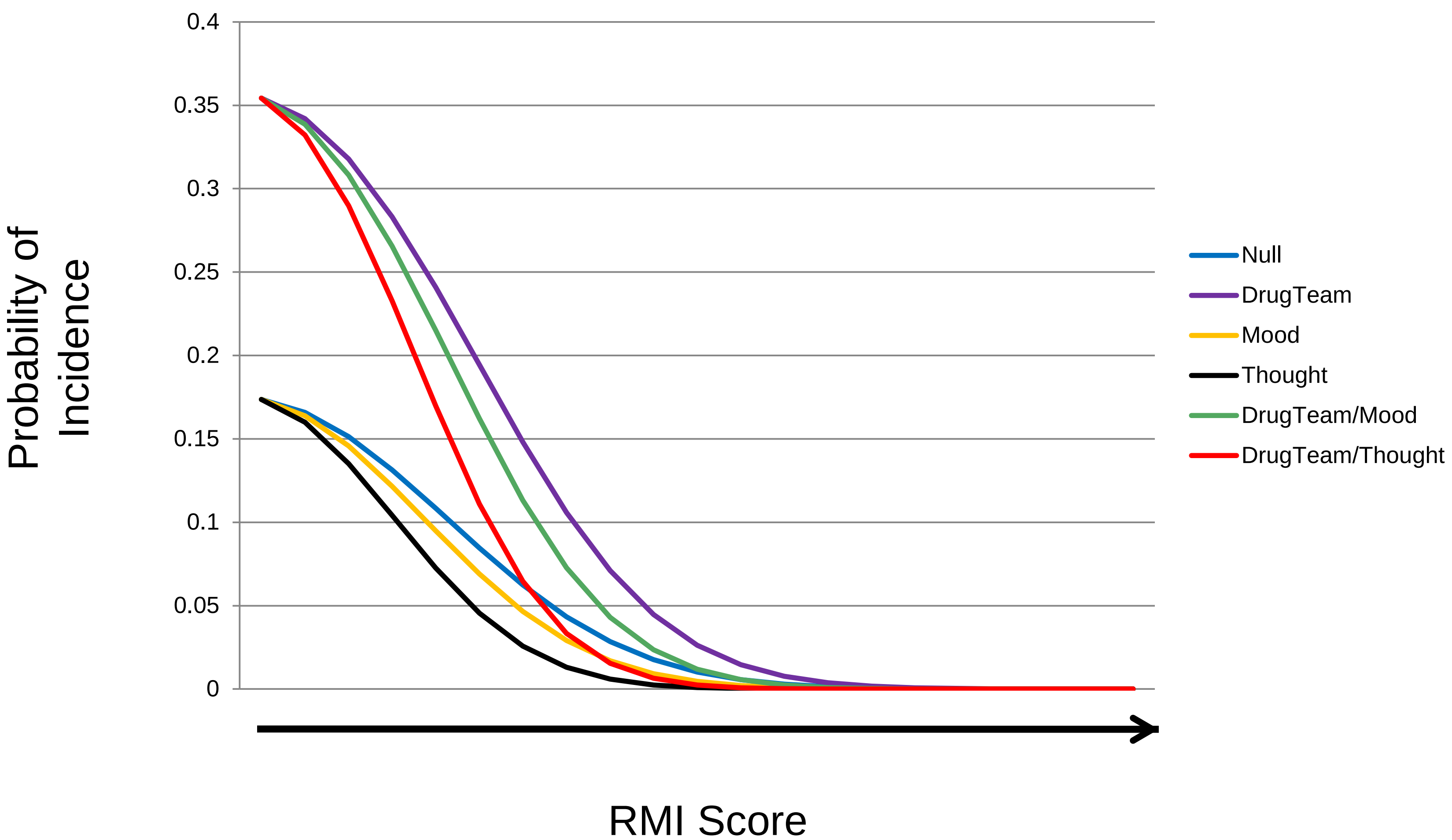
LEVEL 1: $\ln(\pi/1-\pi) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}\,\textbf{(RMIScore)} + \boldsymbol{\varepsilon}$

LEVEL 2: $\beta_0 = \boldsymbol{\varphi_{00}} + \boldsymbol{\varphi_{01}}\textbf{(DrugTeam)} + \varphi_{02}(\text{Mood}) + \varphi_{03}(\text{Thought}) + r_0$

$\beta_1 = \boldsymbol{\varphi_{10}} + \varphi_{11}(\text{DrugTeam}) + \boldsymbol{\varphi_{12}}\textbf{(Mood)} + \boldsymbol{\varphi_{13}}\textbf{(Thought)} + r_1$

Parameter	ψ00	ψ01	ψ10	ψ12	ψ13
Estimate	-1.569	0.96	-0.219	-0.059	-0.176
SE	0.136	0.305	0.036	0.022	0.023
p-value	<0.001	0.03	<0.001	0.009	<0.001

Results



Probability of Incidence
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Null
DrugTeam
Mood
Thought
DrugTeam/Mood
DrugTeam/Thought
RMI Score

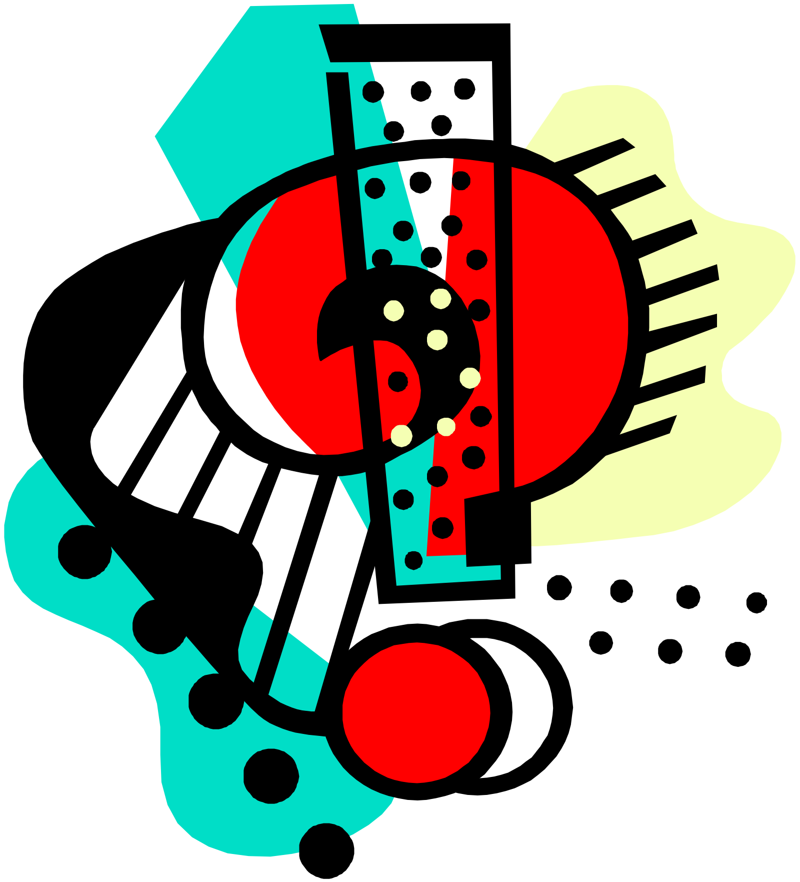
Conclusions

1. As expected there was an overall increase in the prevalence of substance abuse in the drug treatment teams, but a significant difference in slope was not found.

2. One item of interest is that the rate of decrease was dependant on whether the consumer had a mood or thought disorder. Both were higher than the case of a general disorder, but thought was much greater. It is thought that this results from the substance abuse being more directly related to the symptoms of thought disorders, thus as recovery increases it would be more likely for those to stop the substance abuse.

3. Overall, the negative relationship with the recovery marker score indicates that an increase in recovery supports also helps reduce the prevalence of substance abuse.

Questions???



Misuses of HLM

1. Data is not Hierarchical

    - Since HLM has become so popular, the incidence of researchers using HLM for non-hierarchical data has also increased.

2. The variance estimates are not significantly different from zero at the higher levels.

    - If the all variance components of the higher level effects is 0, this implies these are fixed/constant in the lower levels, thus HLM is not needed.

Misuses of HLM

LEVEL 1: $\text{Score} = \mathbf{Grade_{intercept}} + \mathbf{Grade_{slope}} + \varepsilon_{11}$

LEVEL 2: $\text{Grade}_{\text{intercept}} = \mathbf{SES_{intercept1}} + \mathbf{SES_{slope1}} + r_{21}$

$\text{Grade}_{\text{slope}} = \mathbf{SES_{intercept2}} + \mathbf{SES_{slope2}} + r_{22}$

LEVEL 3: $\text{SES}_{\text{intercept1}} = \mathbf{Public_{intercept1}} + \mathbf{Public_{slope1}} + u_{31}$

$\text{SES}_{\text{slope1}} = \mathbf{Public_{intercept2}} + \mathbf{Public_{slope2}} + u_{32}$

$\text{SES}_{\text{intercept2}} = \mathbf{Public_{intercept3}} + \mathbf{Public_{slope3}} + u_{33}$

$\text{SES}_{\text{slope2}} = \mathbf{Public_{intercept4}} + \mathbf{Public_{slope4}} + u_{34}$

Misuses of HLM

3. Violations of Assumptions

   - It is assumed that the error terms at all levels are equal across units and normally distributed.
   - This can be difficult to assess and if not meet will result in inaccurate inferences.

Overall:

As with all statistical models, HLM models have various assumptions, and violations of these assumptions can result in inference errors and/or utilization of a more complicated than necessary model.

Discussion

References

Books:

•(Basic) Luke, D.A. (2004) Multilevel Modeling, Sage Publications Inc., Thousand Oaks, CA.

•(Advanced) Raudenbush, S.W. & Bryk, A.S. (2002) Hierarchical Linear Models: Applications and Data Analysis Methods , Sage Publications Inc., Thousand Oaks, CA.

Papers:

•Hoffman, D. (1997). An Overview of the Logic and Rationale of Hierarchical Linear Models. *Journal of Management*, 23(6), 723-744.

•Streenbergen, M & Jones, B. (2002) Modeling Multilevel Data Structures. *American Journal of Political Science*, 46(1), 218-237

HLM in SPSS:

•Peugh, J. & Enders, C. (2005). Using the SPSS Mixed Procedure to Fit Cross-Sectional and Longitudinal Multilevel Models. *Educational and Psychological Measurement*, 65(5), 717-741.

Consideration in use of HLM:

•Raudenbush, S (1995) Reexamining, Reaffirming, and Improving Application of Hierarchical Linear Models. *Journal of Educational and Behavioral Statistics, 20(2), 210-220*

•Leeuw, J. & Kreft, I. (1995). Questioning Multilevel Models. *Journal of Educational and Behavioral Statistics, 20(2), 210-220*