Searching for Latent Patterns in Longitudinal Data: How Well Do Latent Growth Mixture Models Work?

Kevin L. Delucchi, PhD
University of California, San Francisco

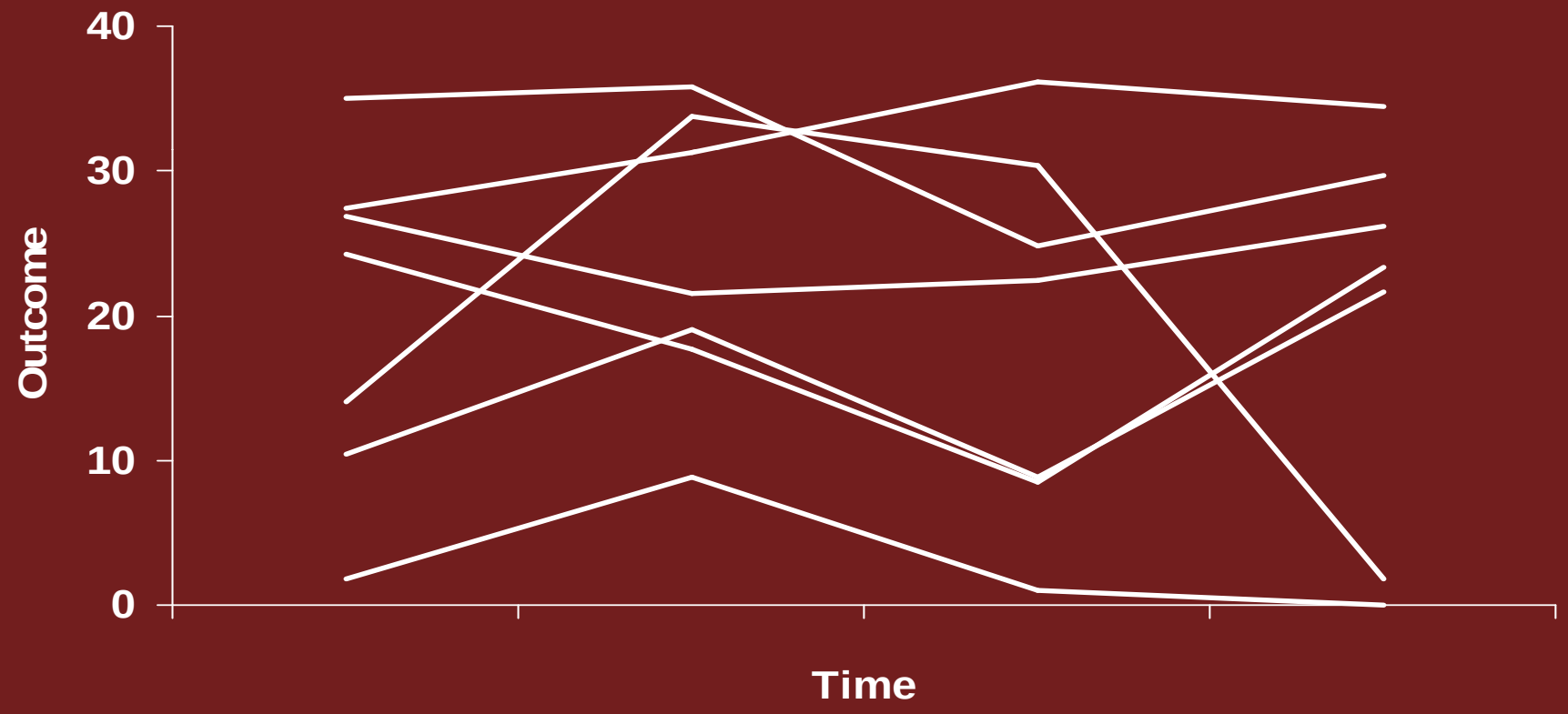
Acknowledgements

- Alan Bostrom, Ph.D.
- Connie Weisner, DPH
- Alcohol Research Group
- Sharon Hall, Ph.D.
- SF Treatment Research Center

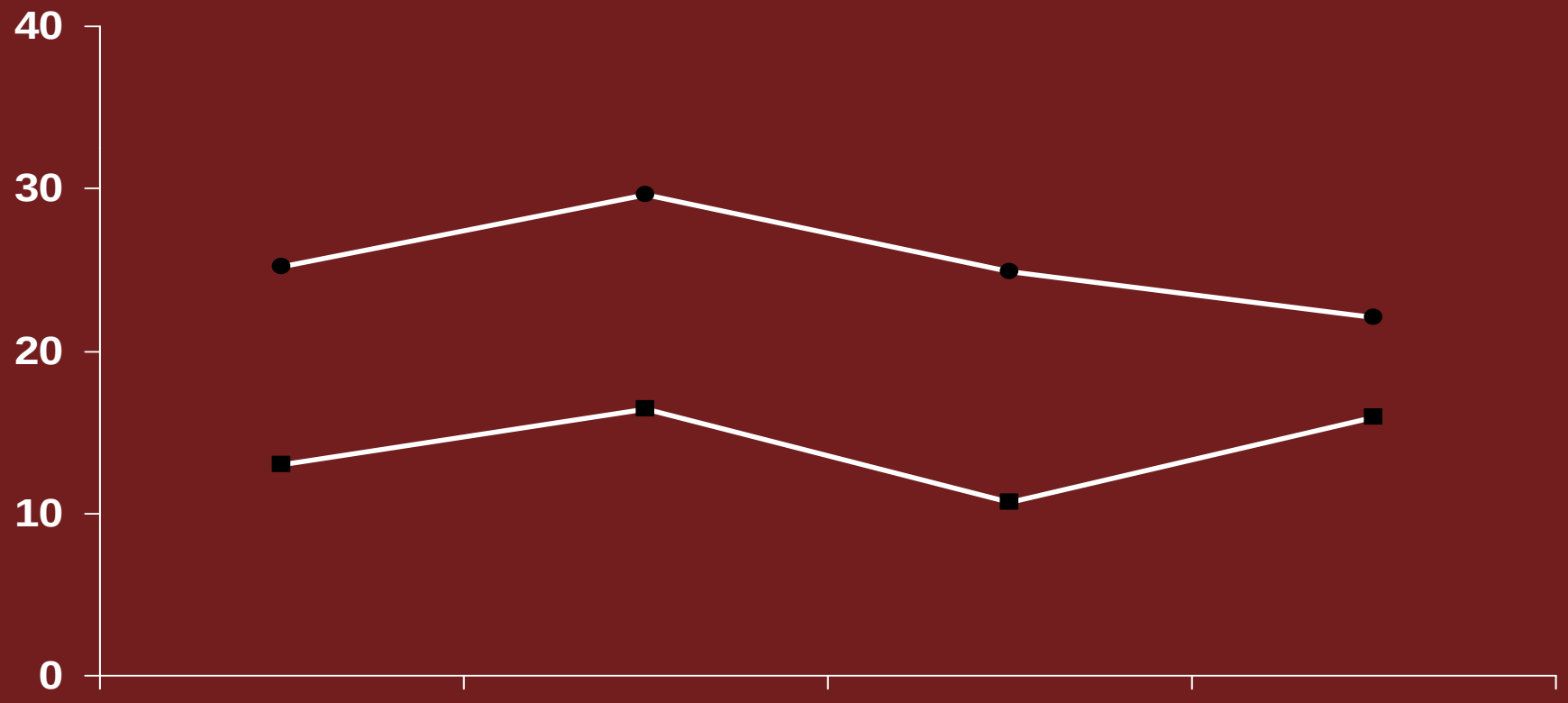
Latent Constructs

- Common in social sciences
- Depression, affiliation, social pressure, big-five personality dimensions
- Factor analysis and structural equation modeling
- Growth mixture modeling

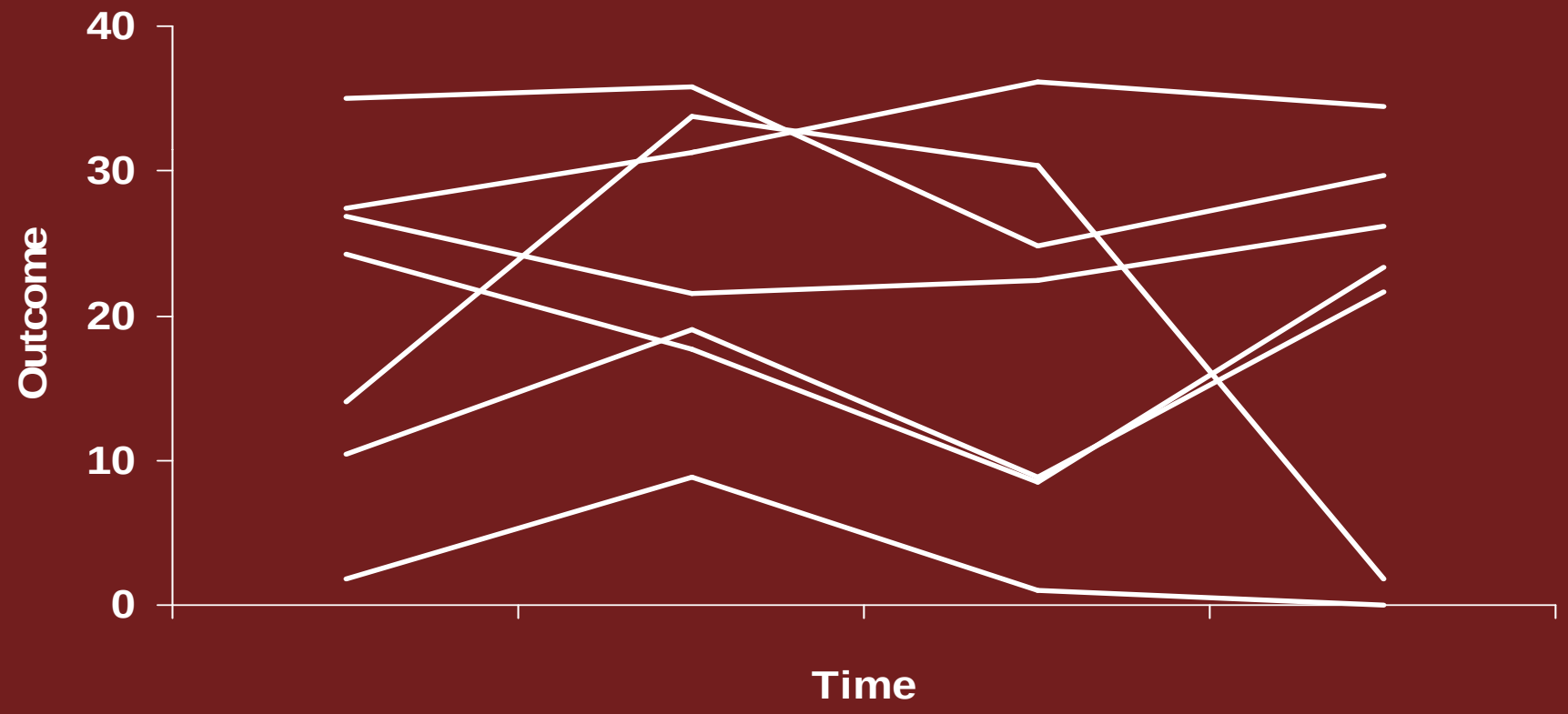
Individual Trajectories



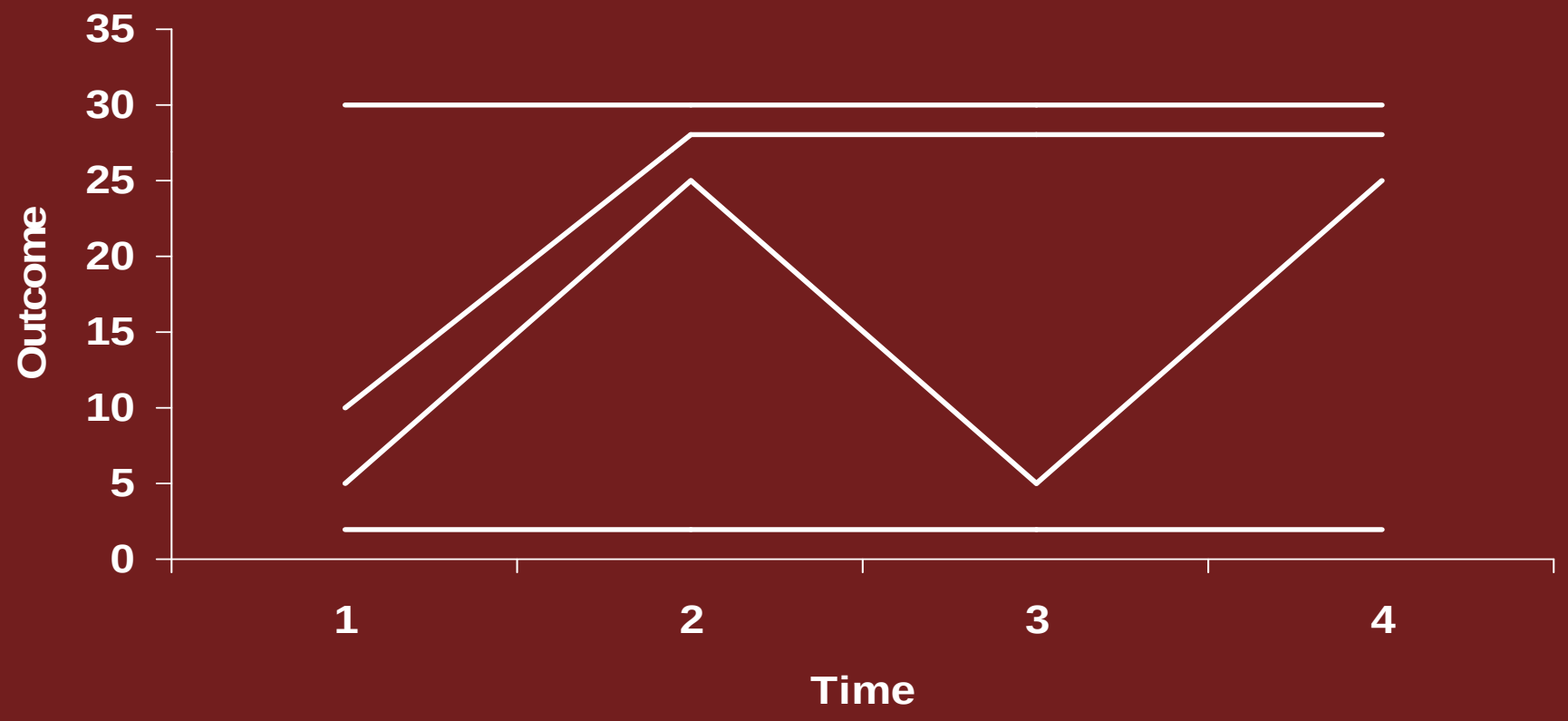
Known Group Trajectories



Individual Trajectories



Latent Trajectories



Part 1:
Motivating Example

Trajectories of Alcohol Consumption over 5 Years

Delucchi, K. L., Matzger, H., & Weisner, C. (2004). Dependent and problem drinking over five years: A latent class growth analysis. *Drug and Alcohol Dependence, 74*, 235-244.

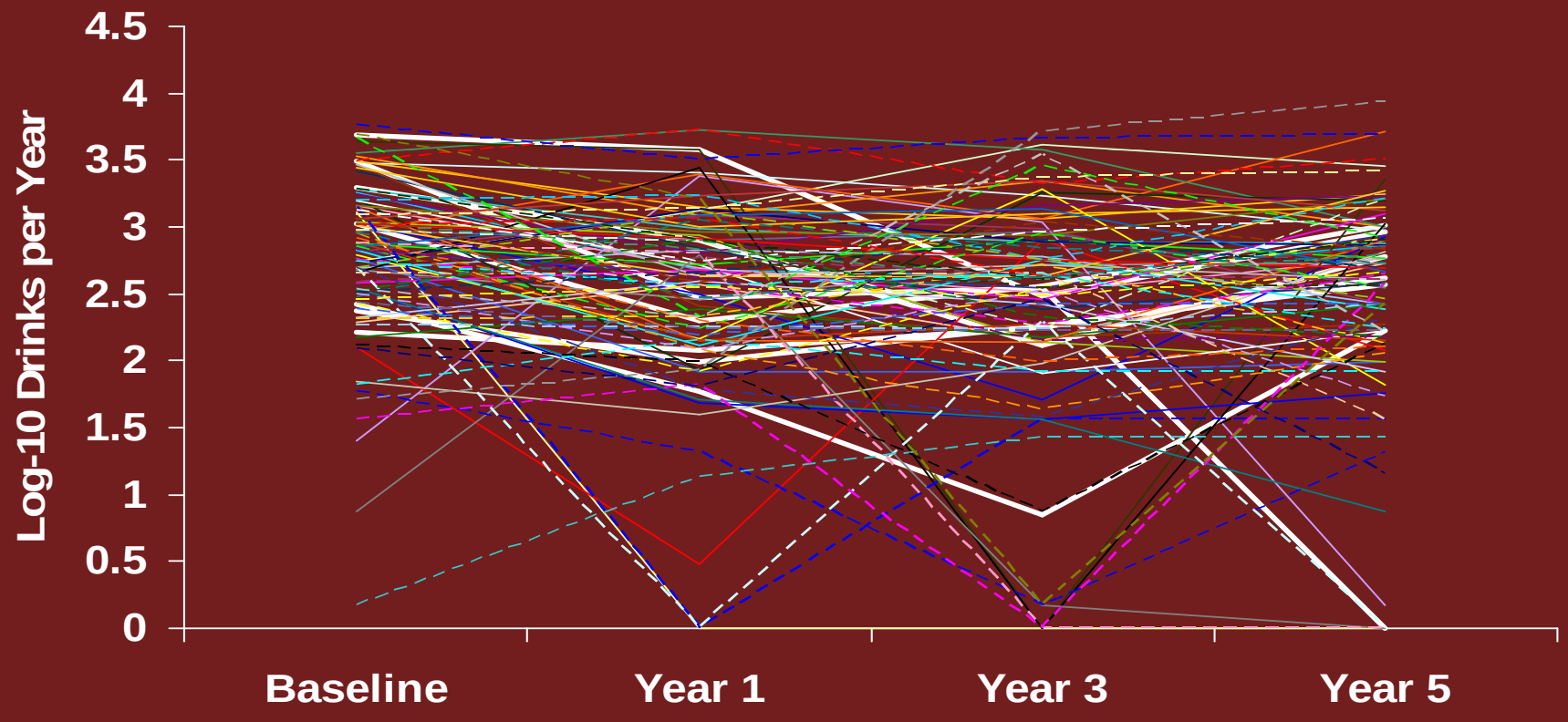
Research Goals

- Understanding the long-term course of problematic drinking
- Common patterns of drinking?
- Covariates related to those patterns?

Study Design

- Longitudinal survey of dependent and problem alcohol drinkers
- Data from 5 years (4 assessments)
- N = 1094 (complete data)
- Outcome: N drinks per year

First 100 Cases



Known Groups

- Mixed-effects models
  - Hierarchical linear models
  - Random effects
- Estimate effects for sex, treatment condition, etc.

Growth Mixture Models
Latent Class Growth Models

- Random effects models in K subgroups
- Finite mixture modeling
- Set of observed trajectories → smaller set of latent trajectories
- Improvement on "classify-then-analyze"
- Measures of model fit for model selection
- Software: Proc Traj, MPlus

Outcome Modeled

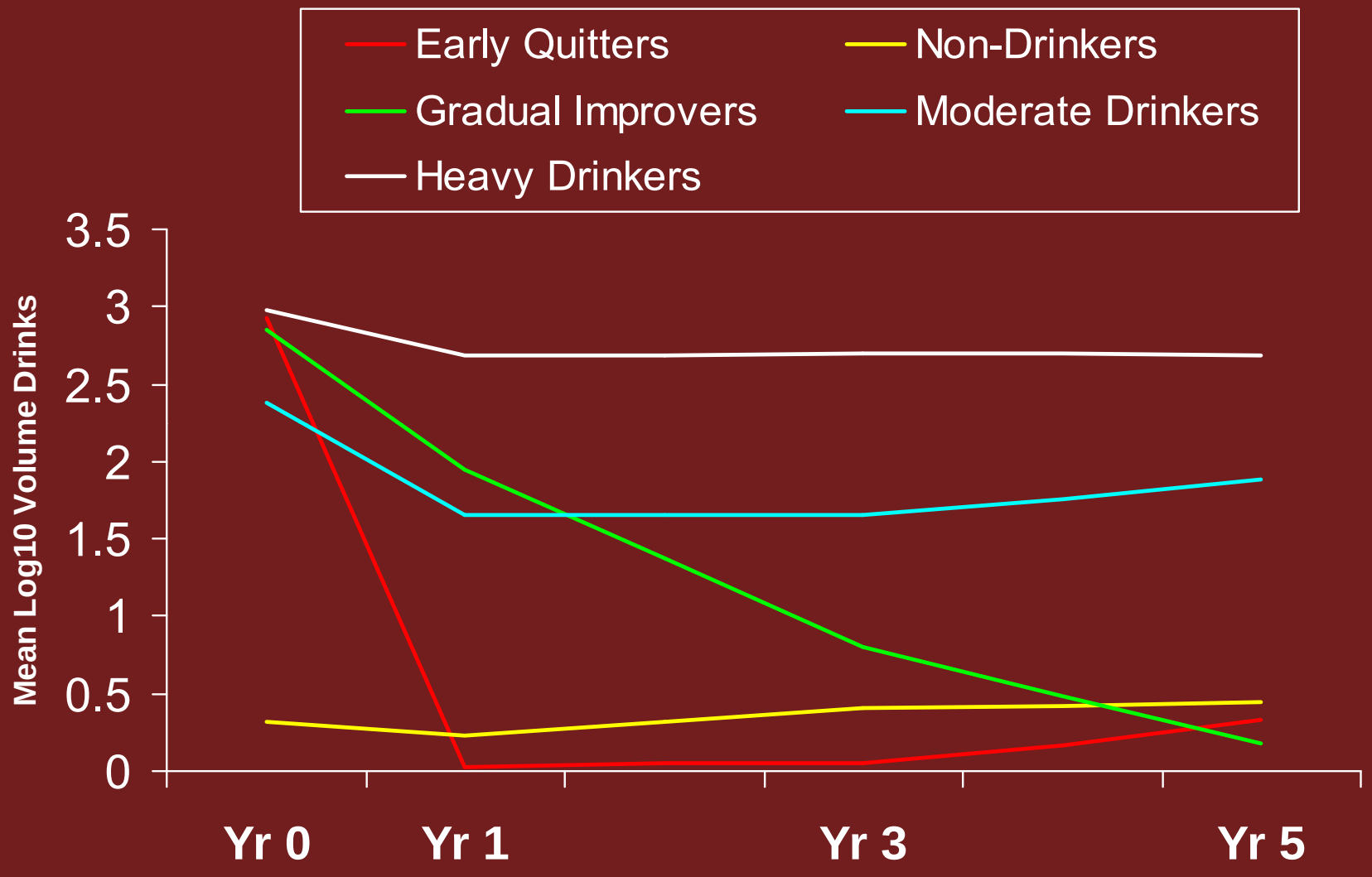
- Number of drinks of alcohol in prior year
- Log-10 transformed
- Fit models with 2 to 6 latent groupings
- Added covariates

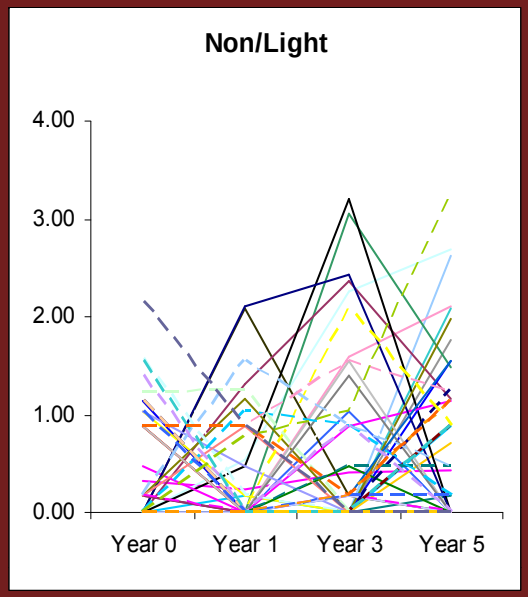
Number of Classes	BIC
2	-6558.3
3	-6432.5
4	-6328.7
5	-6261.7
6	-6496.1

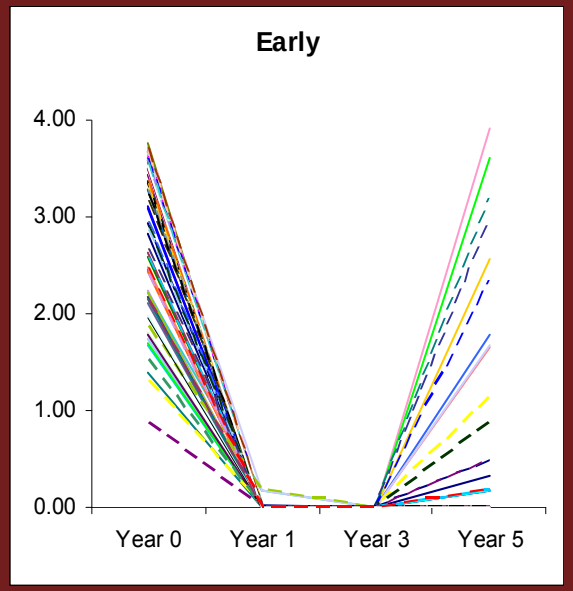
Results

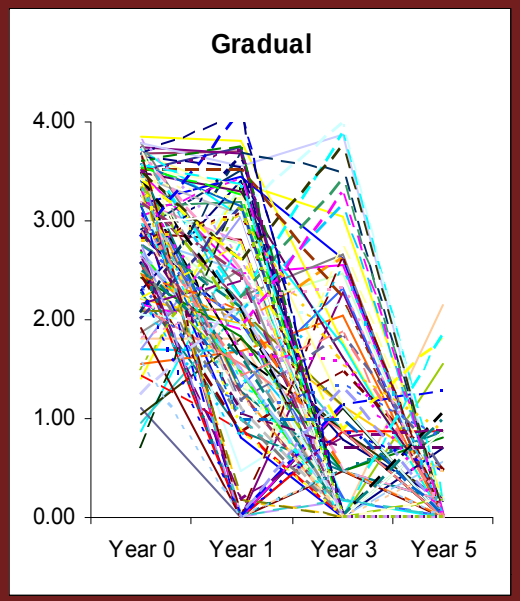
Five class model produced best fit

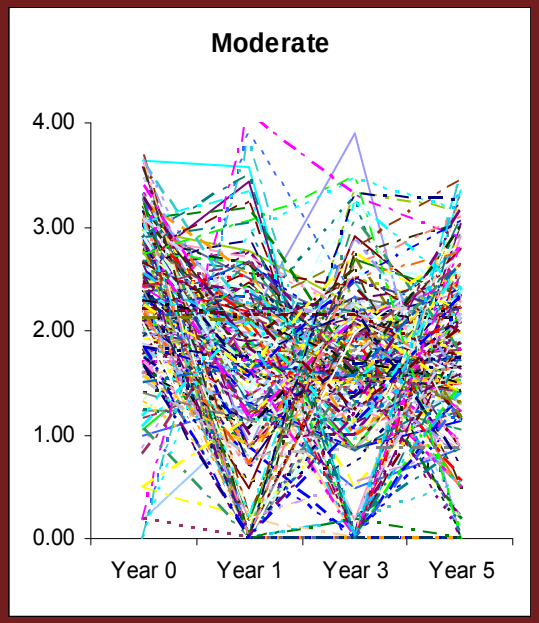
- Early Quitters (N=88)
- Light/Non-drinkers (N=76)
- Gradual Improvers (N=129)
- Moderate Drinkers (N=229)
- Heavy Drinkers (N=572)

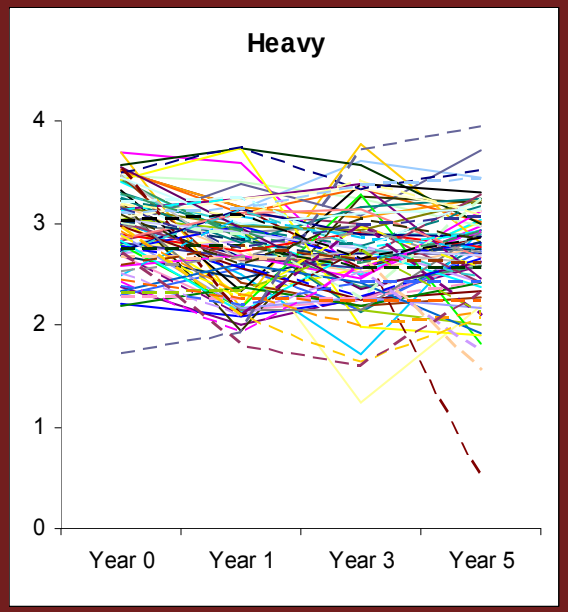

Early Quitters
Non-Drinkers
Gradual Improvers
Moderate Drinkers
Heavy Drinkers
Mean Log10 Volume Drinks
3.5
3
2.5
2
1.5
1
0.5
0
Yr 0
Yr 1
Yr 3
Yr 5

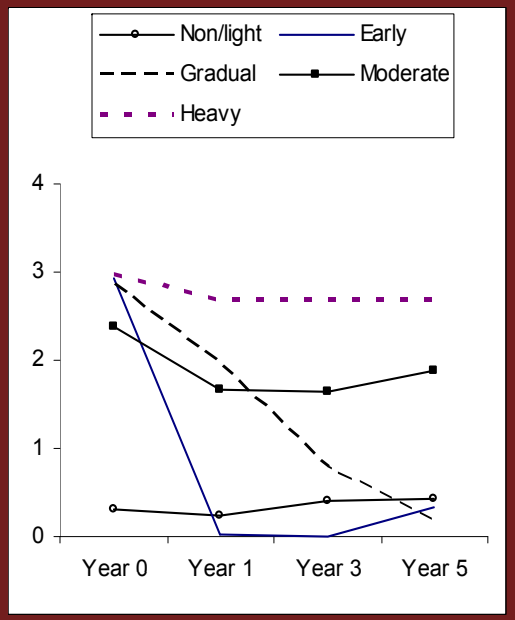

Non/Light
4.00
3.00
2.00
1.00
0.00
Year 0
Year 1
Year 3
Year 5


Early
4.00
3.00
2.00
1.00
0.00
Year 0
Year 1
Year 3
Year 5


Gradual
4.00
3.00
2.00
1.00
0.00
Year 0
Year 1
Year 3
Year 5


Moderate
4.00
3.00
2.00
1.00
0.00
Year 0
Year 1
Year 3
Year 5


Heavy
4
3
2
1
0
Year 0
Year 1
Year 3
Year 5

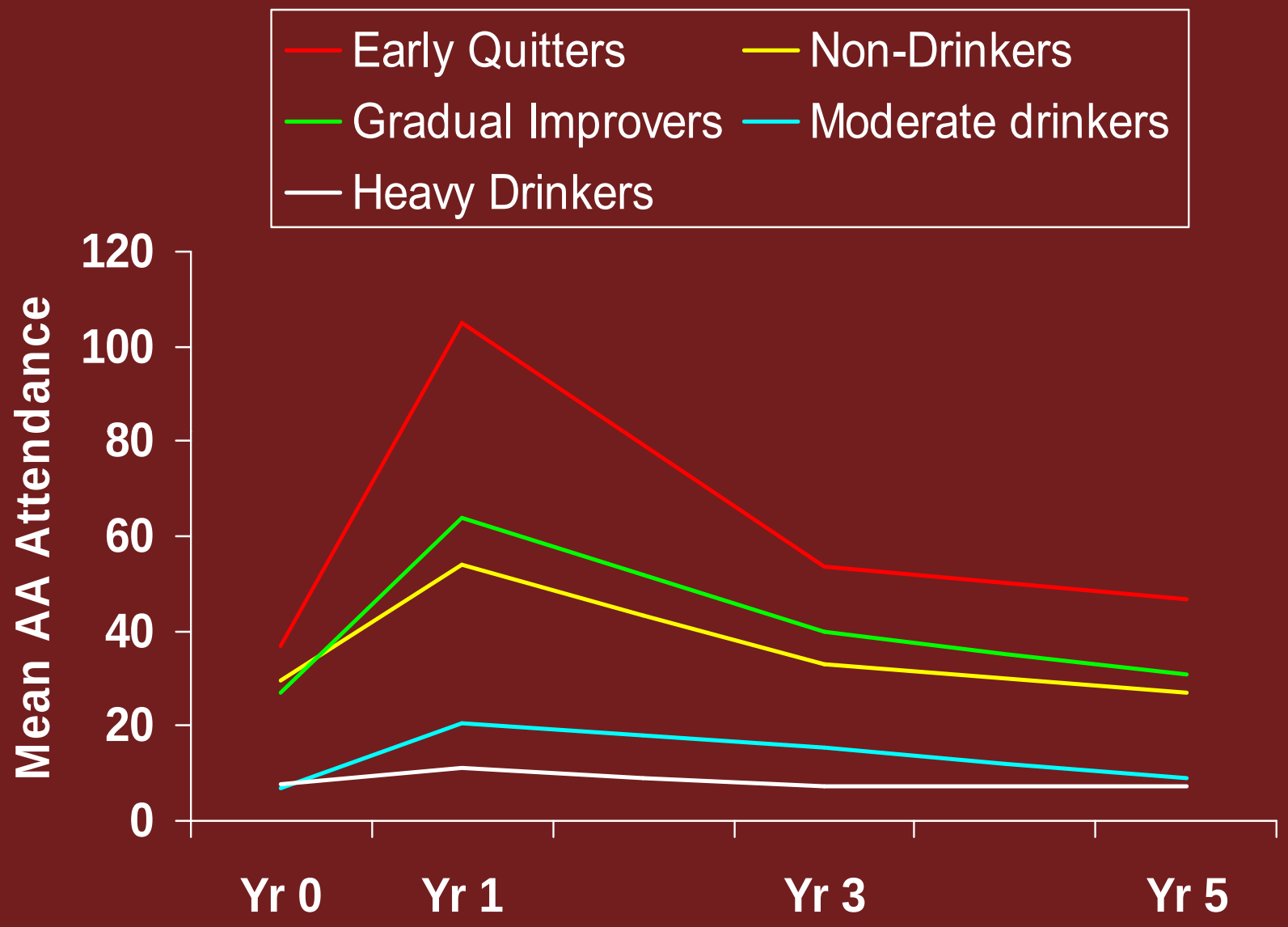

Non/light
Early
Gradual
Moderate
Heavy
4
3
2
1
0
Year 0
Year 1
Year 3
Year 5

Mean Posterior Probabilities

	Early Quit	Non-Drinkers	Gradual Improve	Mod. Drinkers	Heavy Drinkers
Group 1	**.96**	.01	.01	.01	.00
Group 2	.00	**.96**	.00	.01	.00
Group 3	.01	.00	**.89**	.02	.00
Group 4	.03	.03	.07	**.82**	.10
Group 5	.00	.00	.03	.15	**.90**

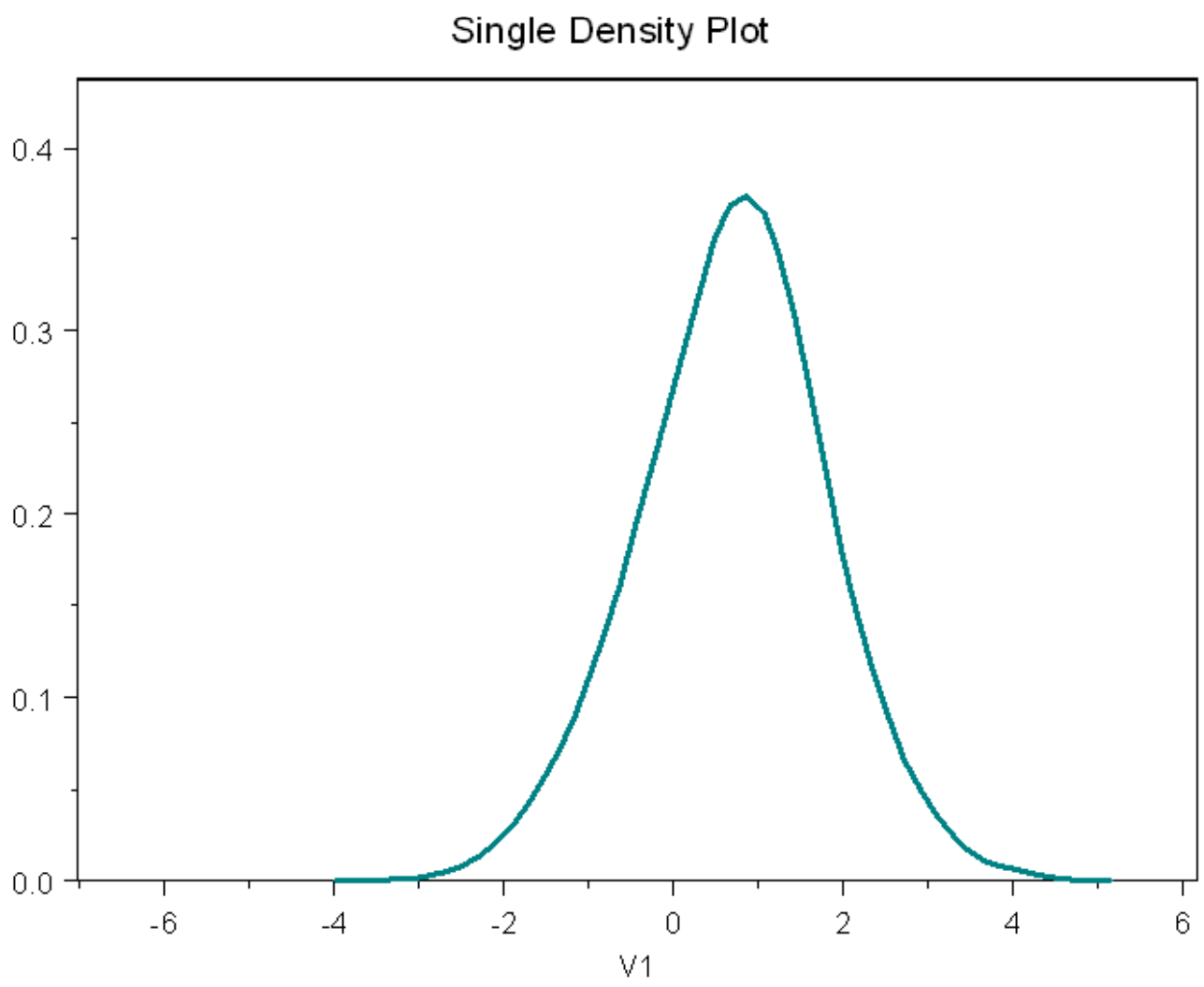
Percent No Drinks Prior Year

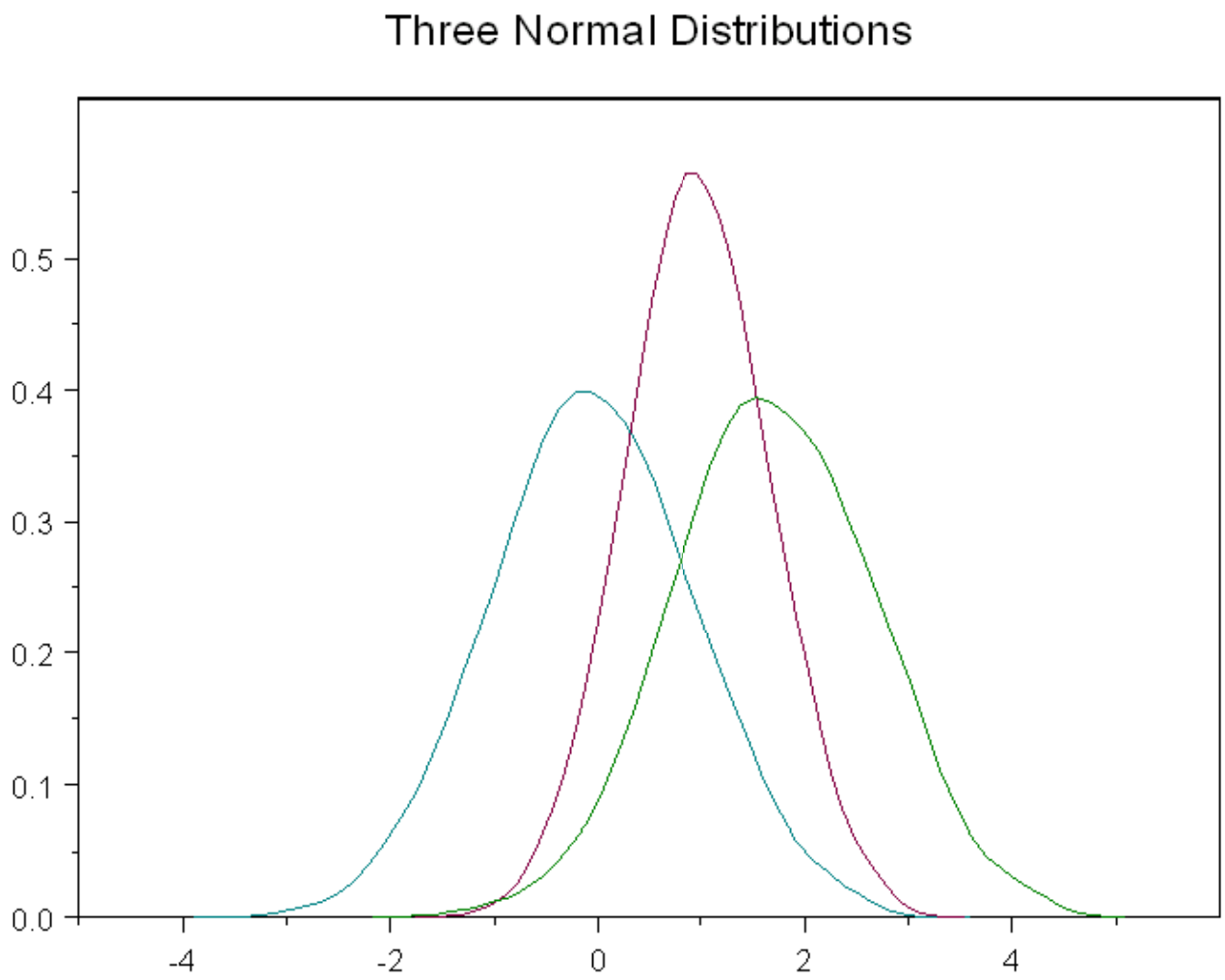
	Baseline	Year 1	Year 3	Year 5
Early Quit	0	97	100	77
Non-Drink	67	76	66	66
Improve	0	13	48	74
Moderate	1	16	13	5
Heavy	0	2	1	1


Early Quitters
Non-Drinkers
Gradual Improvers
Moderate drinkers
Heavy Drinkers
Mean AA Attendance
120
100
80
60
40
20
0
Yr 0
Yr 1
Yr 3
Yr 5

Part 2:

How well do growth mixture models work?


Single Density Plot
0.4
0.3
0.2
0.1
0.0
-6
-4
-2
0
2
4
6
V1


Three Normal Distributions
0.5
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4

Growth Mixture Modeling

- Mixture modeling; Newcomb, 1886, Pearson, 1884
- Muthén & Shedden (1999)
- Nagin: latent class mixture modeling

Prior Research

Eggleston, et al. (2004) – length of follow-up effecting trajectory shapes

Jackson & Sher (2006) – effects of number of assessments

Henson, et al. (2007) fit statistics not accurate for small N (< 500)

Nylund, et al. (2007) - determining number of classes

Lubke & Muthén (2007) - FMM poor capture of classes without covariates

Current Project

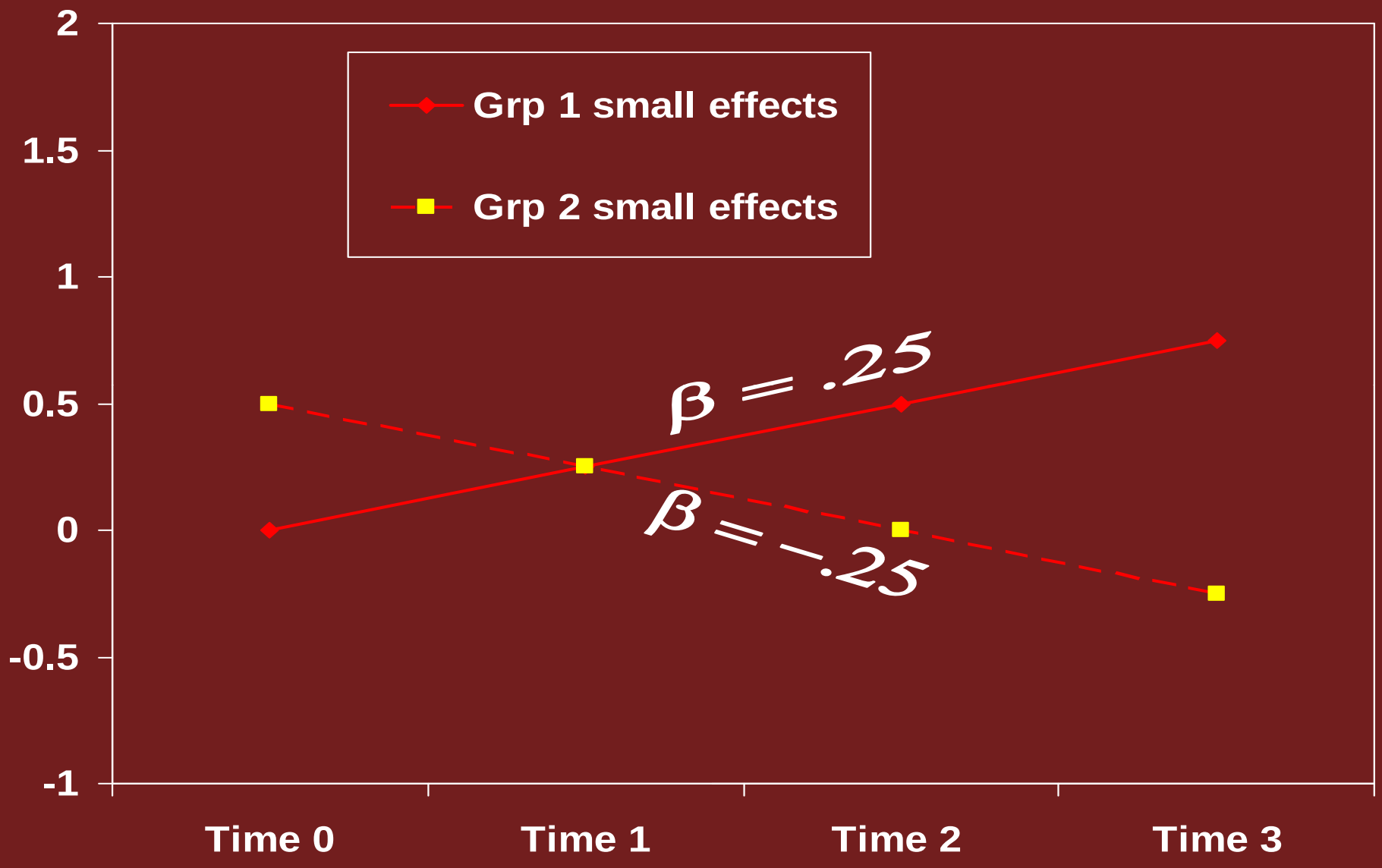
- How well do growth mixture models capture group membership and model parameters?
- Simulations
- Two-group, longitudinal design
- Percent correctly classified, estimates of intercepts and slopes

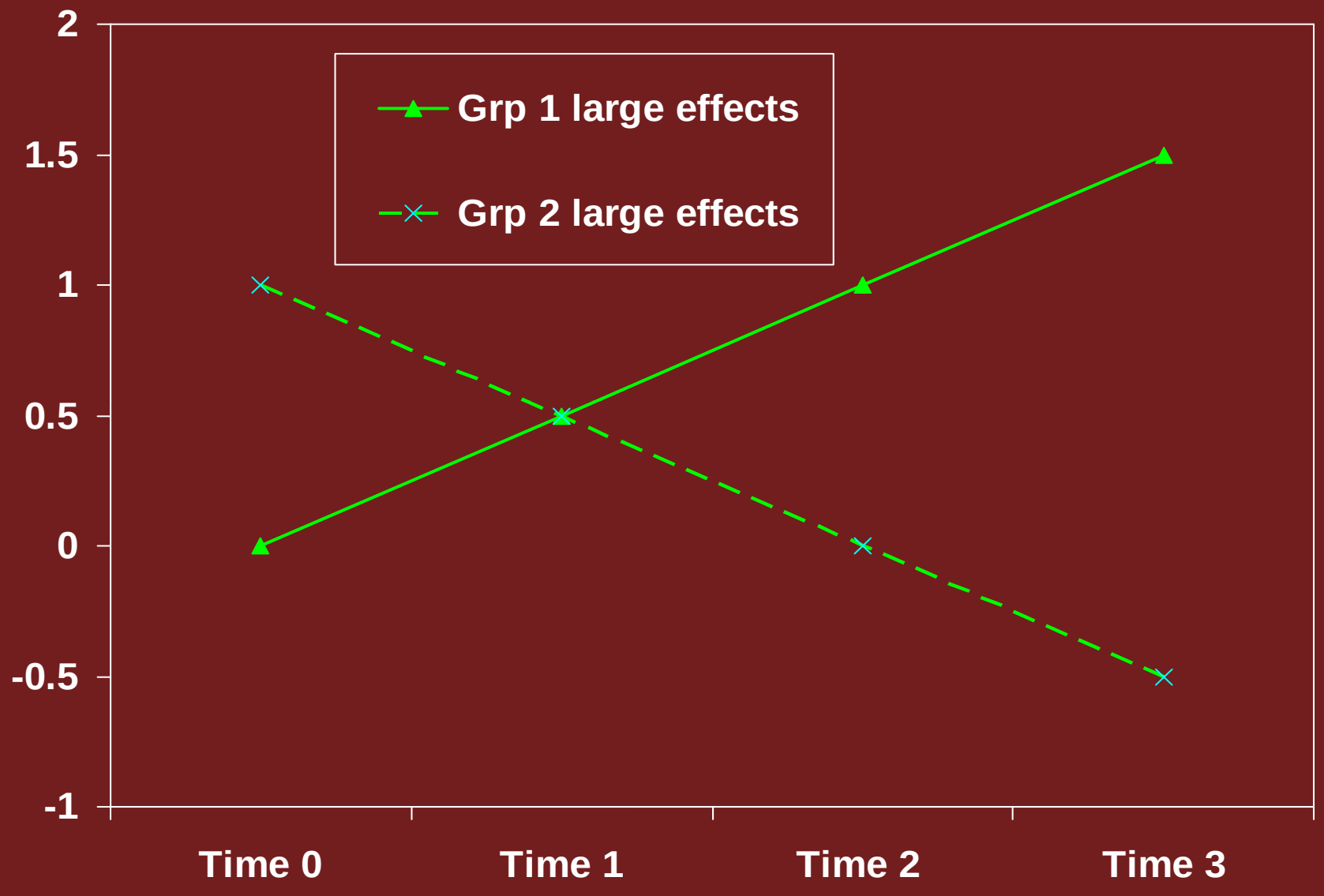
Initial Conditions Simulated

112 cell-design (2x2x2x2x7)

- 2 sample sizes (300 and 900)
- 2 intercept effect sizes
- 2 slope effect sizes
- 2 levels of residual variance
- 7 levels of sample imbalance

- Intercept means -- 2 levels: $(\mu_{I1},\mu_{I2}) = (0,0.5)$ or $(0,1)$
- Slope means -- 2 levels: $(\mu_{S1},\mu_{S2}) = (0.25, -0.25)$ or $(0.5, -0.5)$
- Residual variance -- 2 levels: $\sigma^2_R = 0.1$ or 0.5.
- Imbalance -- 7 levels: $(N1, N2) = (480,420)$, $(540,360)$, $(600,300)$, $(660,240)$, $(720,180)$, $(780,120)$, $(840,60)$


2
1.5
1
0.5
0
-0.5
-1
Grp 1 small effects
Grp 2 small effects
β = .25
β = -.25
Time 0
Time 1
Time 2
Time 3


2
1.5
1
0.5
0
-0.5
-1
Grp 1 large effects
Grp 2 large effects
Time 0
Time 1
Time 2
Time 3

Data Model

Random slope and intercept selected from $N(\mu_i, \Sigma)$

$$\mu_i = \begin{bmatrix} \mu_{Ii} \\ \mu_{Si} \end{bmatrix} \qquad \sum \begin{bmatrix} \sigma^2{}_I & \sigma_{IS} \\ \sigma_{IS} & \sigma^2{}_S \end{bmatrix}$$

After sampling, 4 data points generated;

$$Y_i = I + (i-1)S + R_i \qquad \text{for } i = 1,\ldots,4$$

where R_i is $\sigma_R \varepsilon_i$ with s^2_R being a constant residual variance and $e_i \sim N(0,1)$.

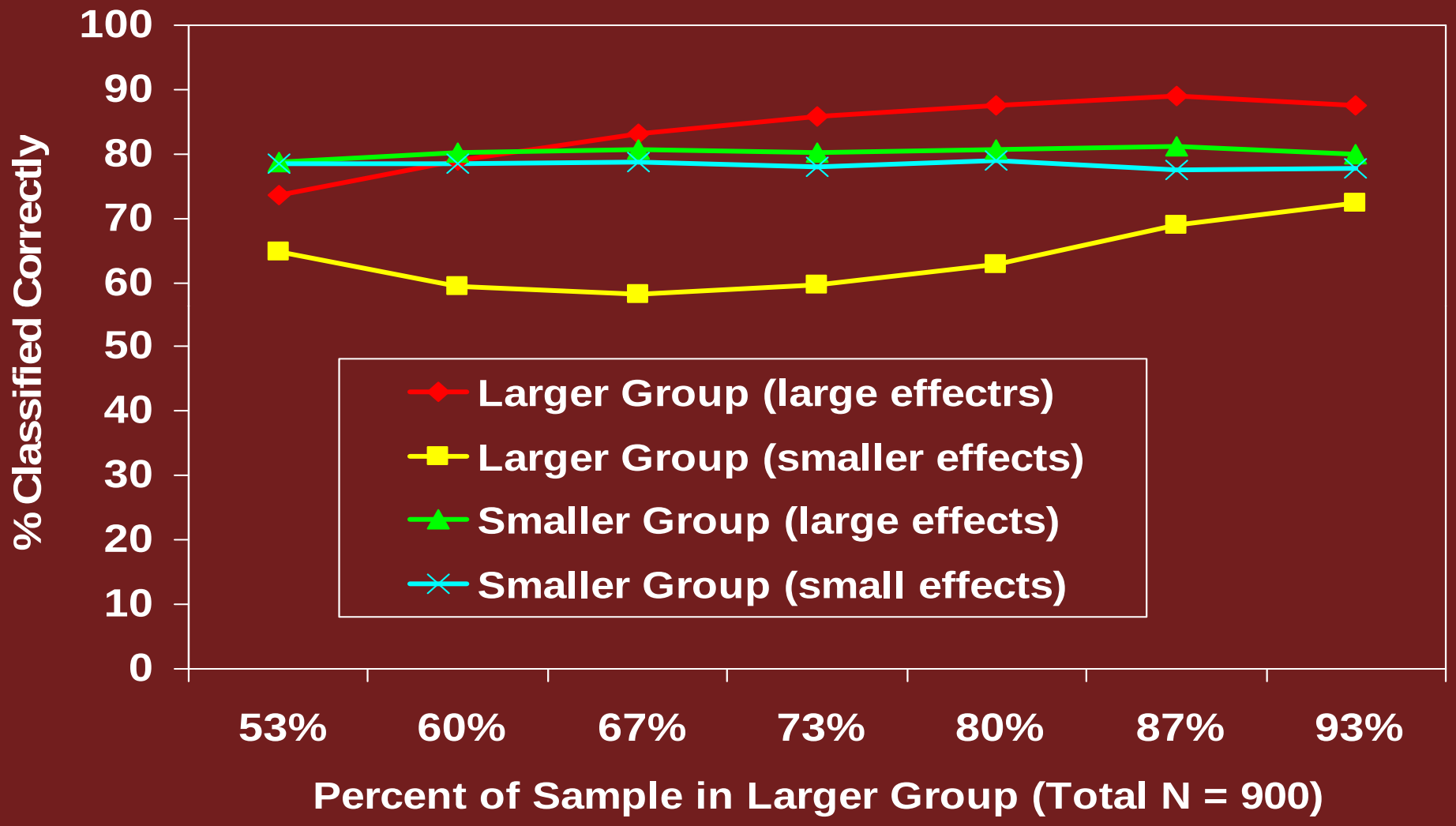
- Data generated in SAS
- Analysis using Mplus (v4.2)
- Data summarized using SAS

```
TITLE:      ANALYSIS 1;
  DATA:       FILE is C:\Delucchi\MPlus\Data\Sample1.txt;
  VARIABLE:   NAMES are Group Subject Y1-Y4;
              CLASSES = C(2);
              USEVARIABLES = Y1-Y4;
  ANALYSIS:   TYPE = MIXTURE;
              STARTS = 100, 10;
  MODEL: %OVERALL%
     i s|y1@0 y2@1 y3@2 Y4@3;
     [c#1] (alpha);
     %c#2%
  SAVEDATA:    FILE = C:\Delucchi\MPlus\Data\File1.txt;
               RESULTS = C:\Delucchi\MPlus\Data\Analysis1.txt;
               SAVE = CPROBABILITIES;
  OUTPUT:
  model constraint: alpha>0;
```

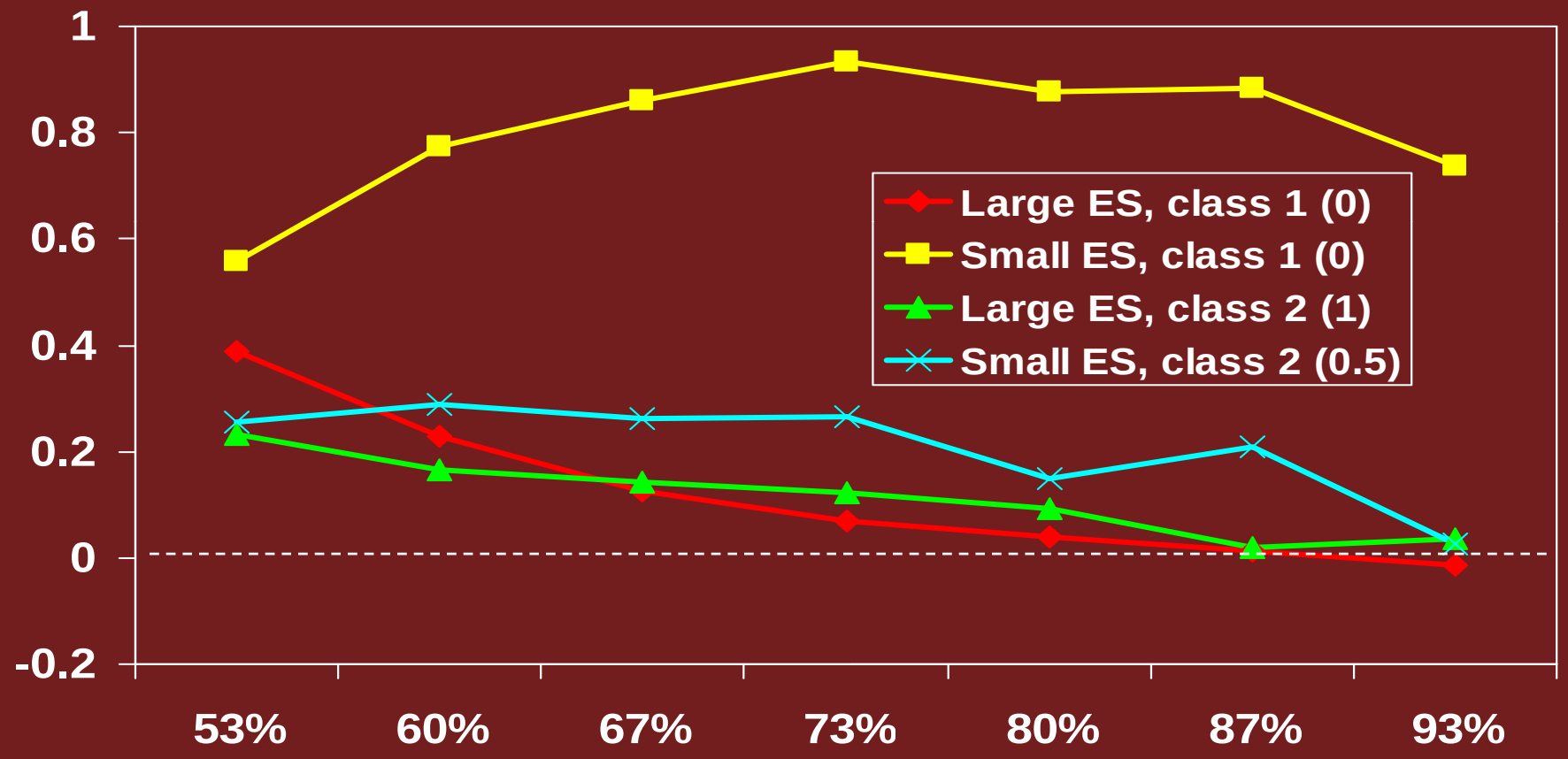
Results from 14 Conditions

- Residual var = 0.5
- N=900
- Largest and smallest effects
- All 7 levels of imbalance

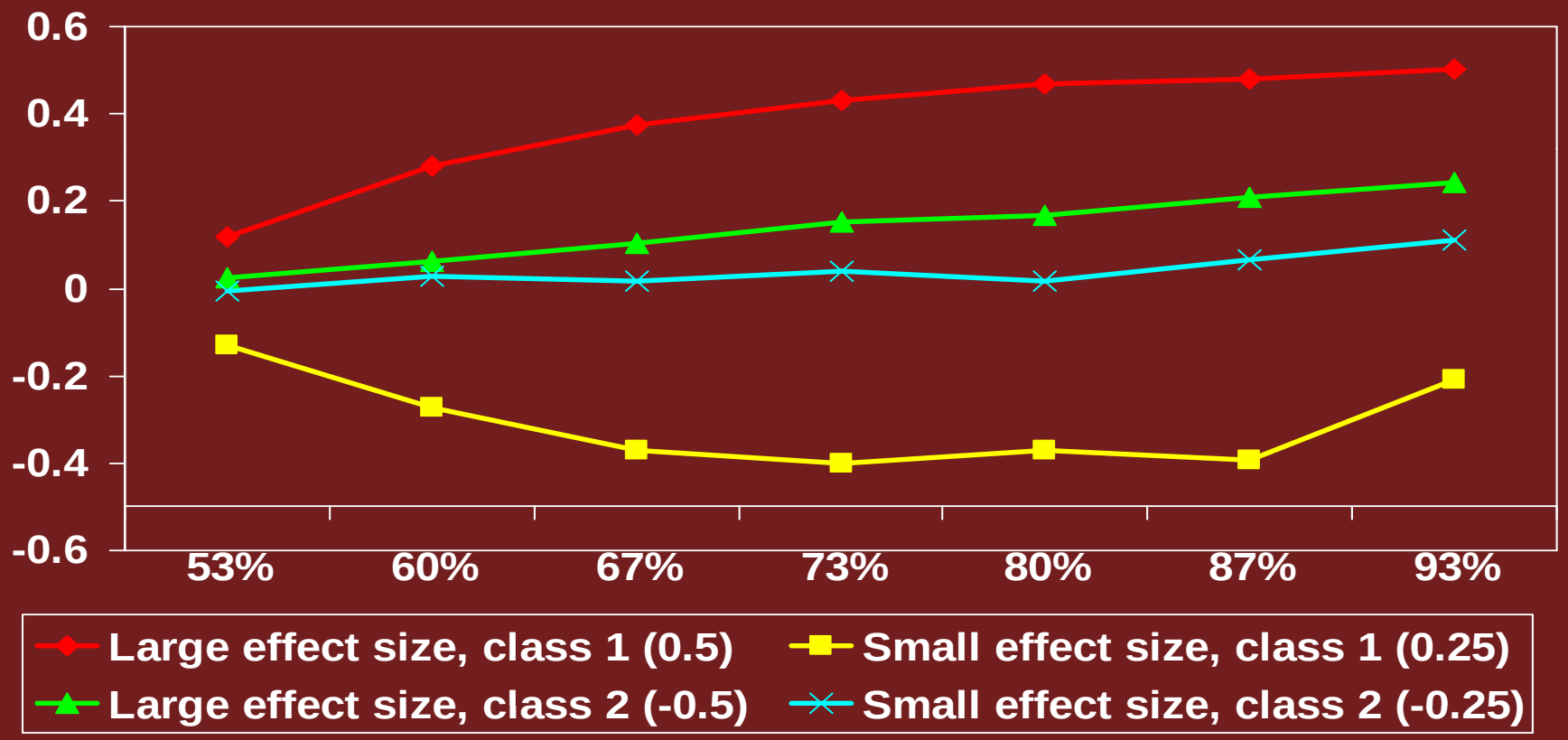
All Four Conditions



Mean Estimated Intercept by Class and Effect Size

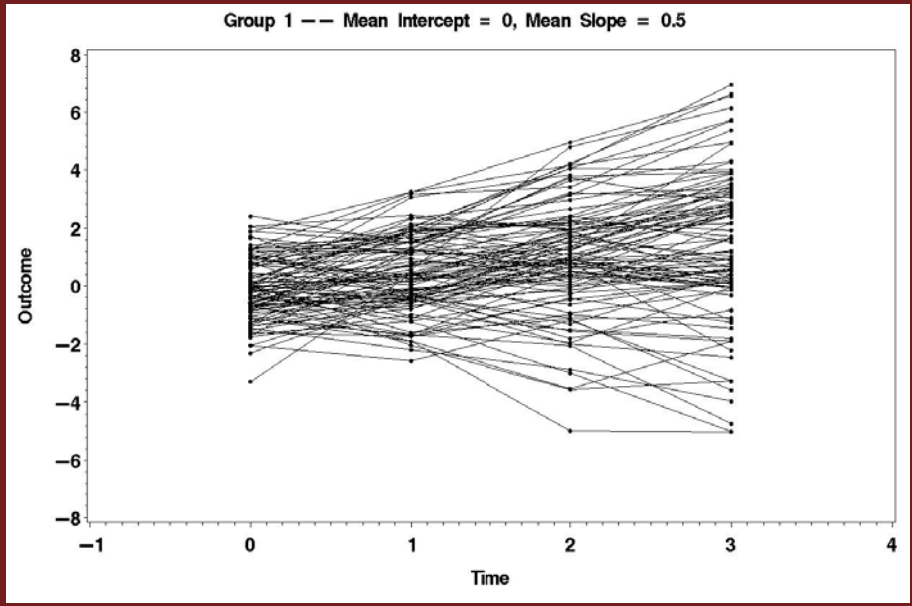


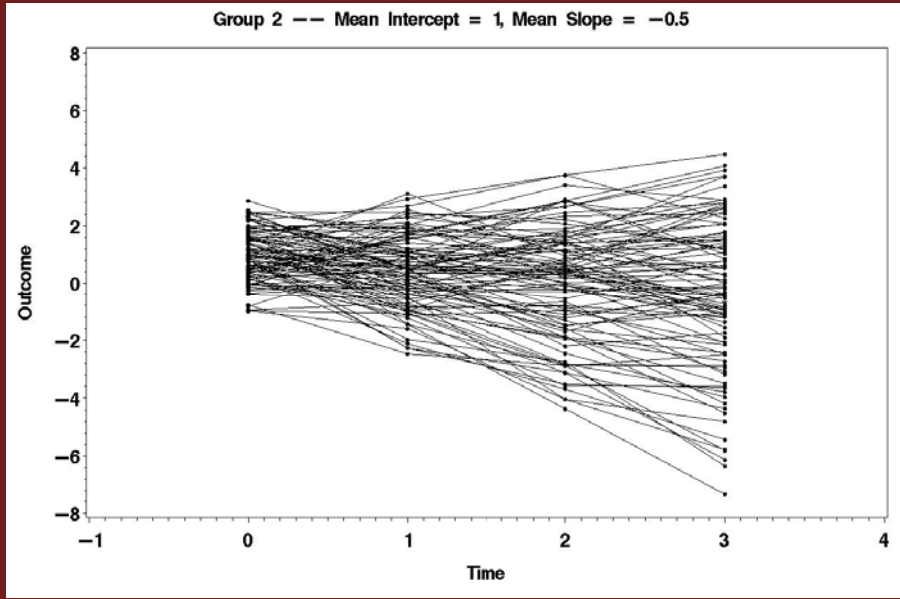
Mean Estimated Slope by Class and Effect Size

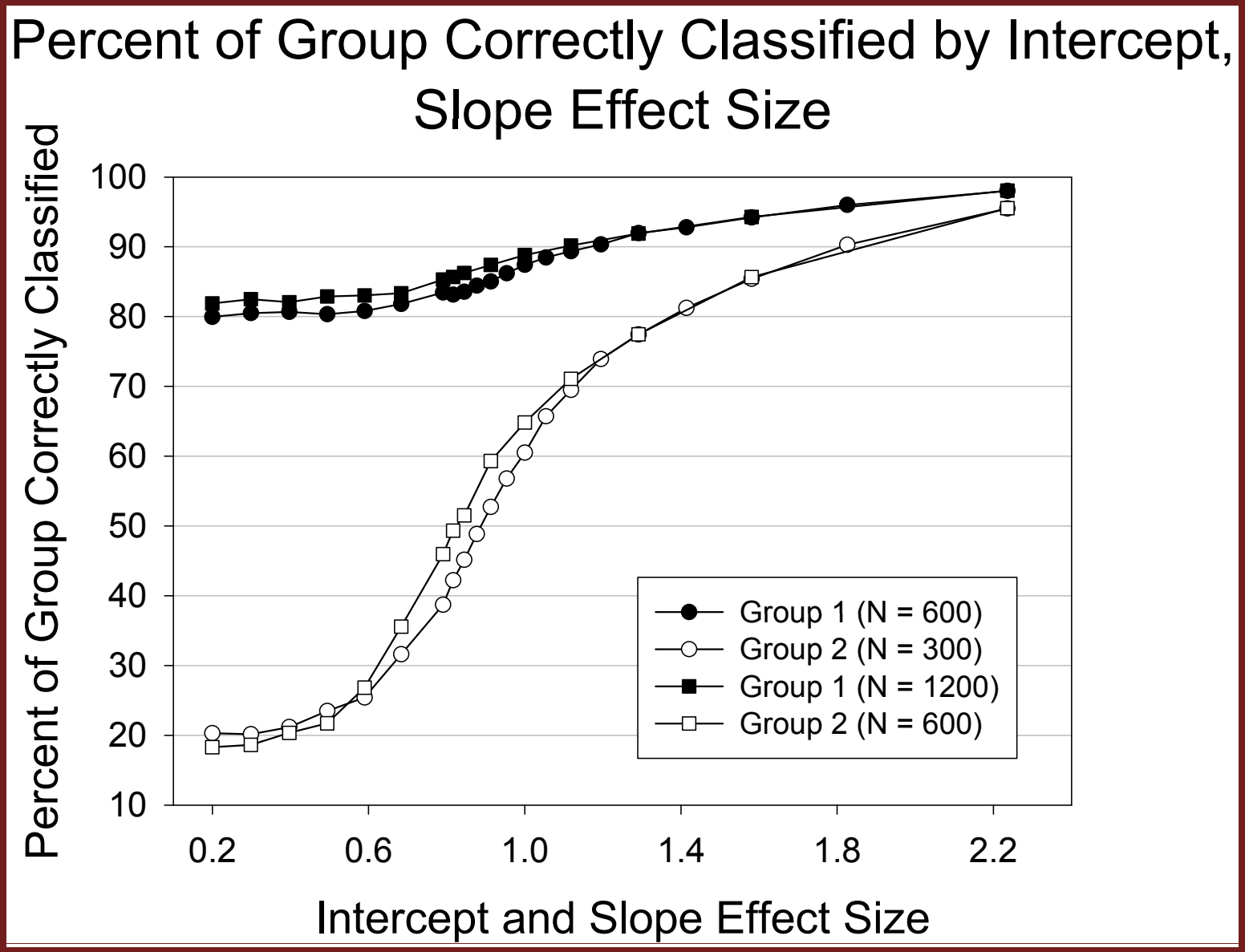


Revised Simulations

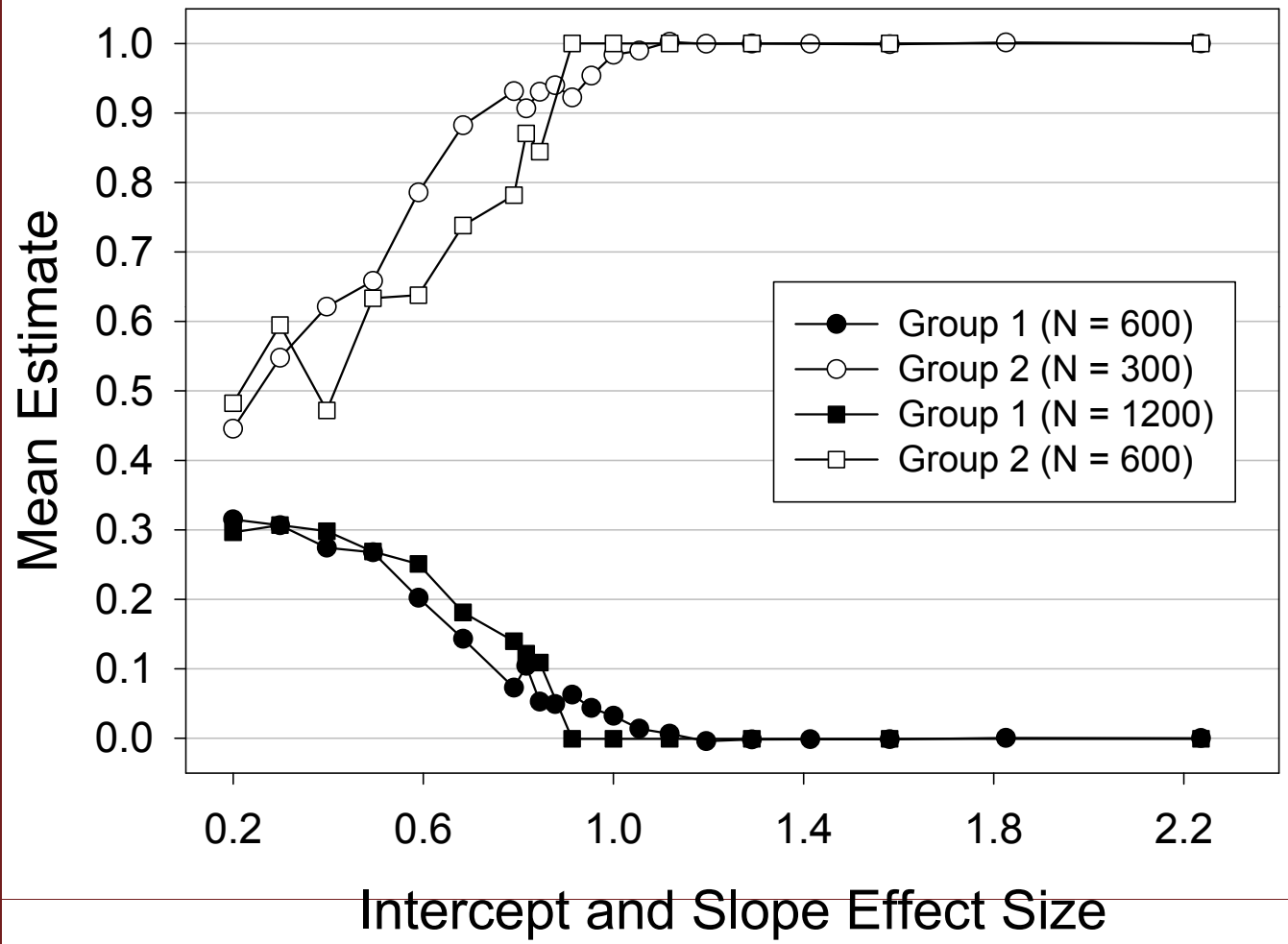
- Ns = 900 (300, 600) and 1800 (600, 1200)
- Residual variance set to 0.1
- Intercepts at 0 and 1
- Slopes at -0.5 and 0.5
- Effect size from 0.2 to 2.2

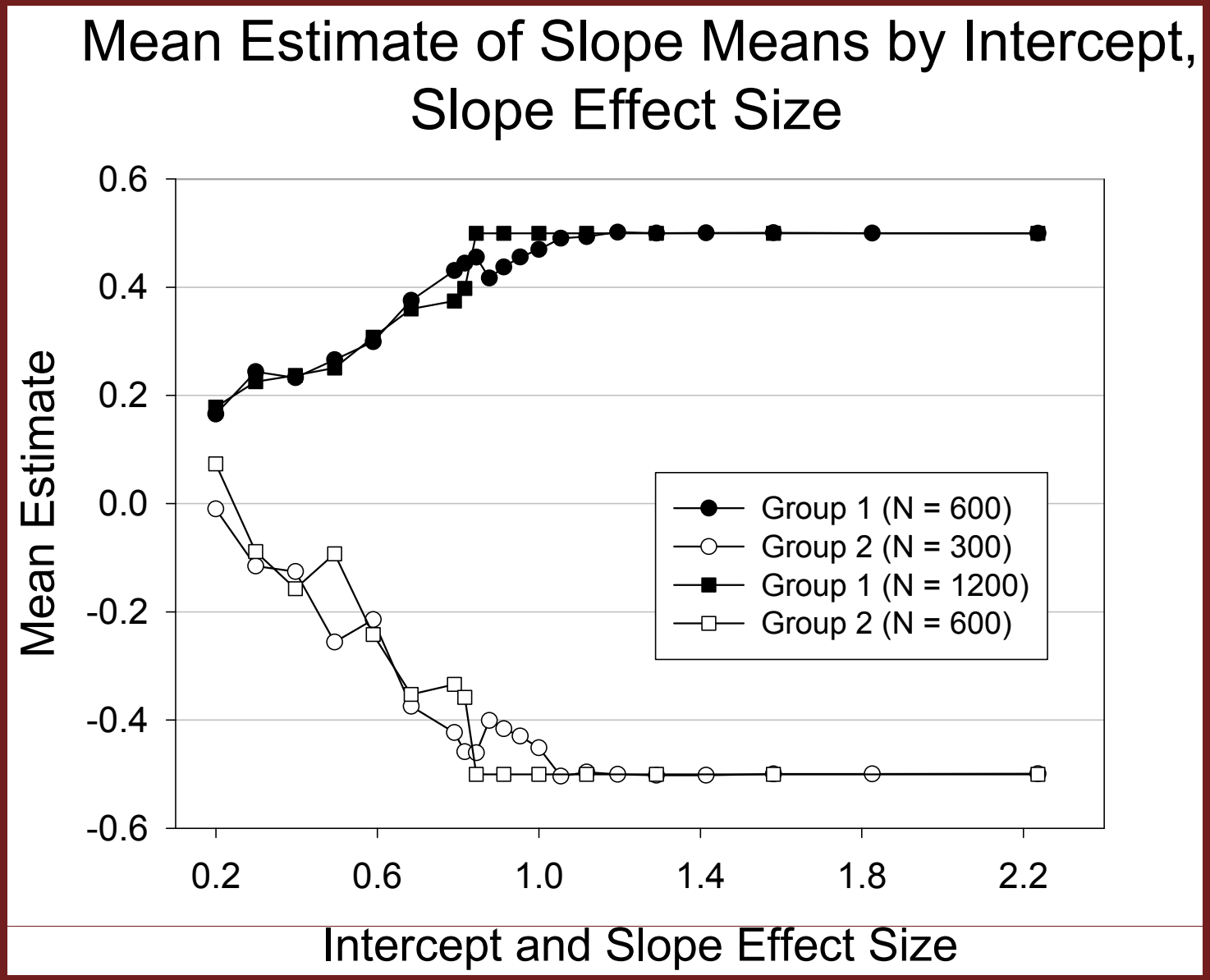

Group 1 -- Mean Intercept = 0, Mean Slope = 0.5
Outcome
Time
8
6
4
2
0
−2
−4
−6
−8
−1
0
1
2
3
4


Group 2 -- Mean Intercept = 1, Mean Slope = −0.5
Outcome
Time
8
6
4
2
0
−2
−4
−6
−8
−1
0
1
2
3
4


Percent of Group Correctly Classified by Intercept, Slope Effect Size
Percent of Group Correctly Classified
100
90
80
70
60
50
40
30
20
10
0.2
0.6
1.0
1.4
1.8
2.2
Intercept and Slope Effect Size
Group 1 (N = 600)
Group 2 (N = 300)
Group 1 (N = 1200)
Group 2 (N = 600)

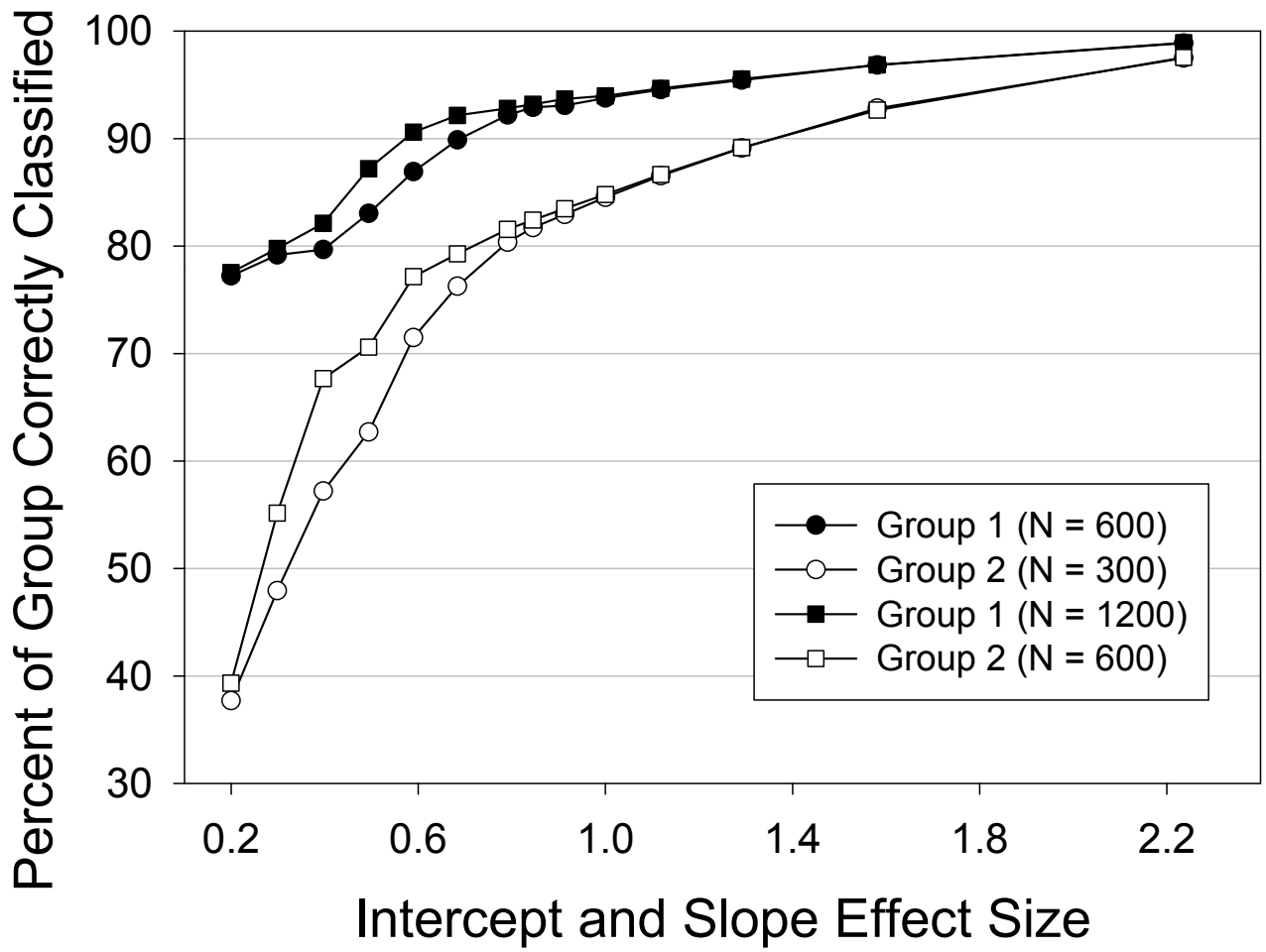
Mean Estimate of Intercept Means by Intercept, Slope Effect Size




Mean Estimate of Slope Means by Intercept, Slope Effect Size
Mean Estimate
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
0.2
0.6
1.0
1.4
1.8
2.2
Group 1 (N = 600)
Group 2 (N = 300)
Group 1 (N = 1200)
Group 2 (N = 600)
Intercept and Slope Effect Size

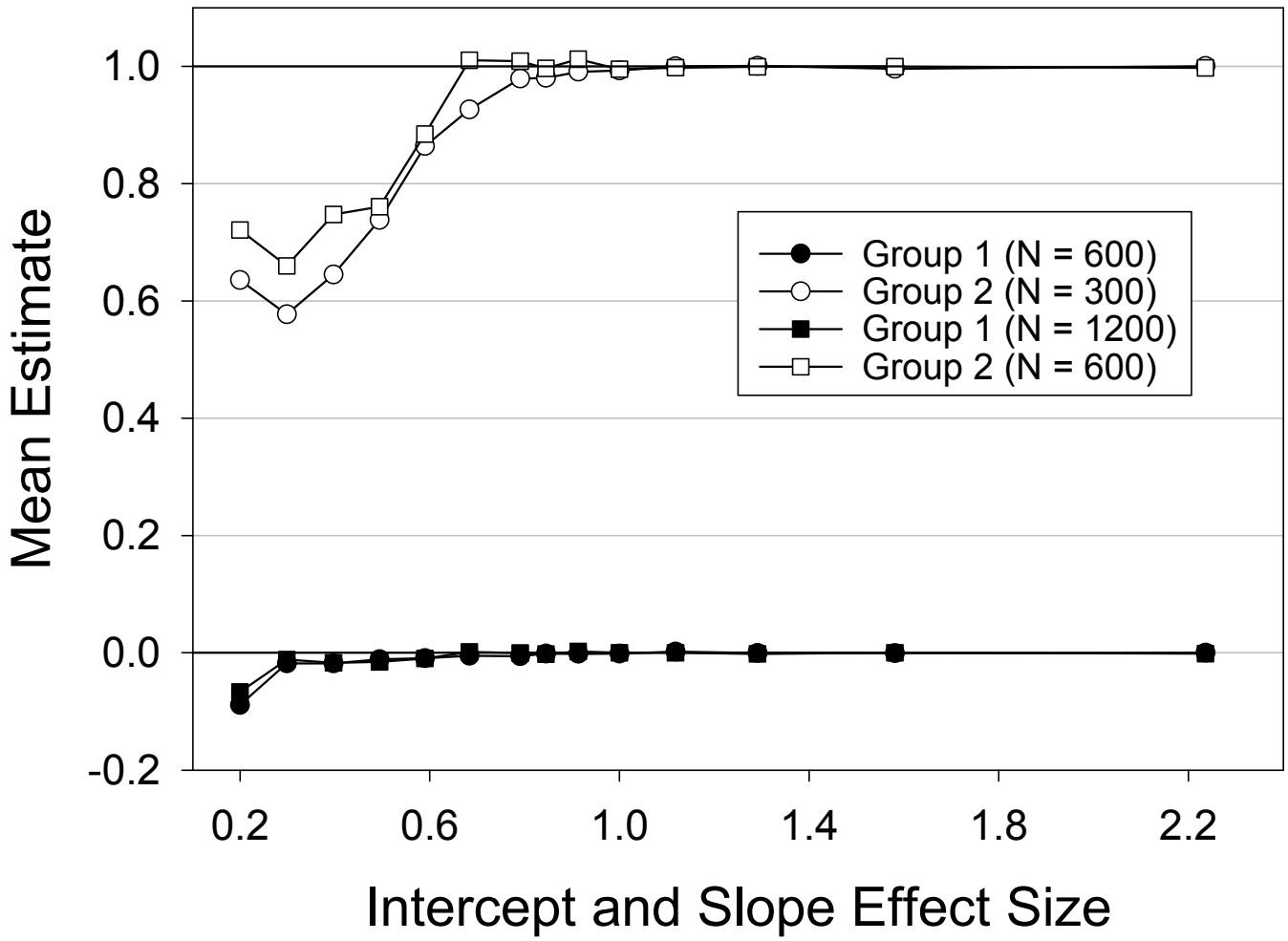
Percent of Group Correctly Classified by Intercept, Slope Effect Size

Residual variance = 0.1, with continuous covariate N(5,5) in Group 1, N(15,5) in Group 2



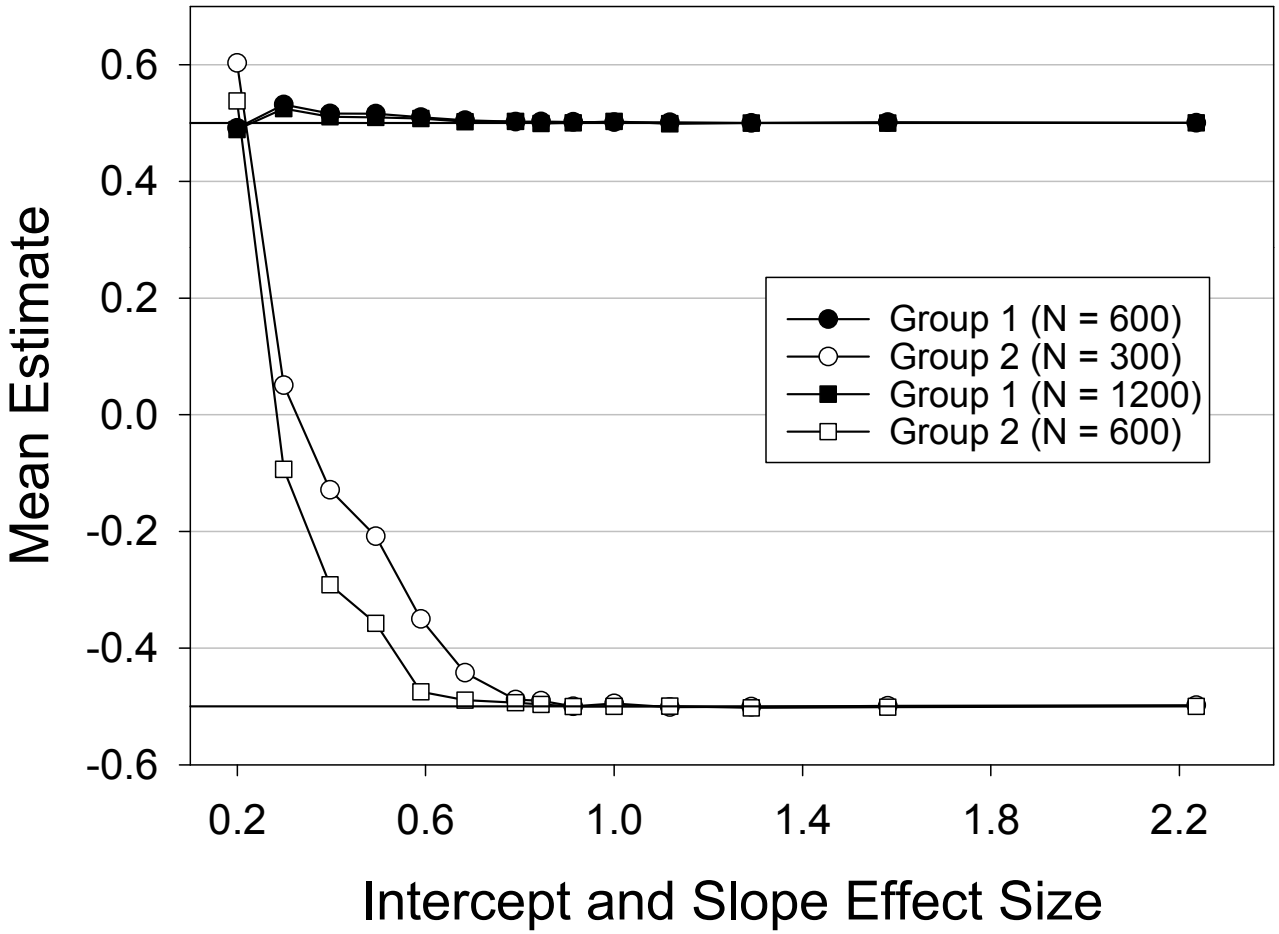
Mean Estimate of Intercept Means by Intercept, Slope Effect Size

Residual variance = 0.1, with continuous covariate N(5,5) in Group 1, N(15,5) in Group 2



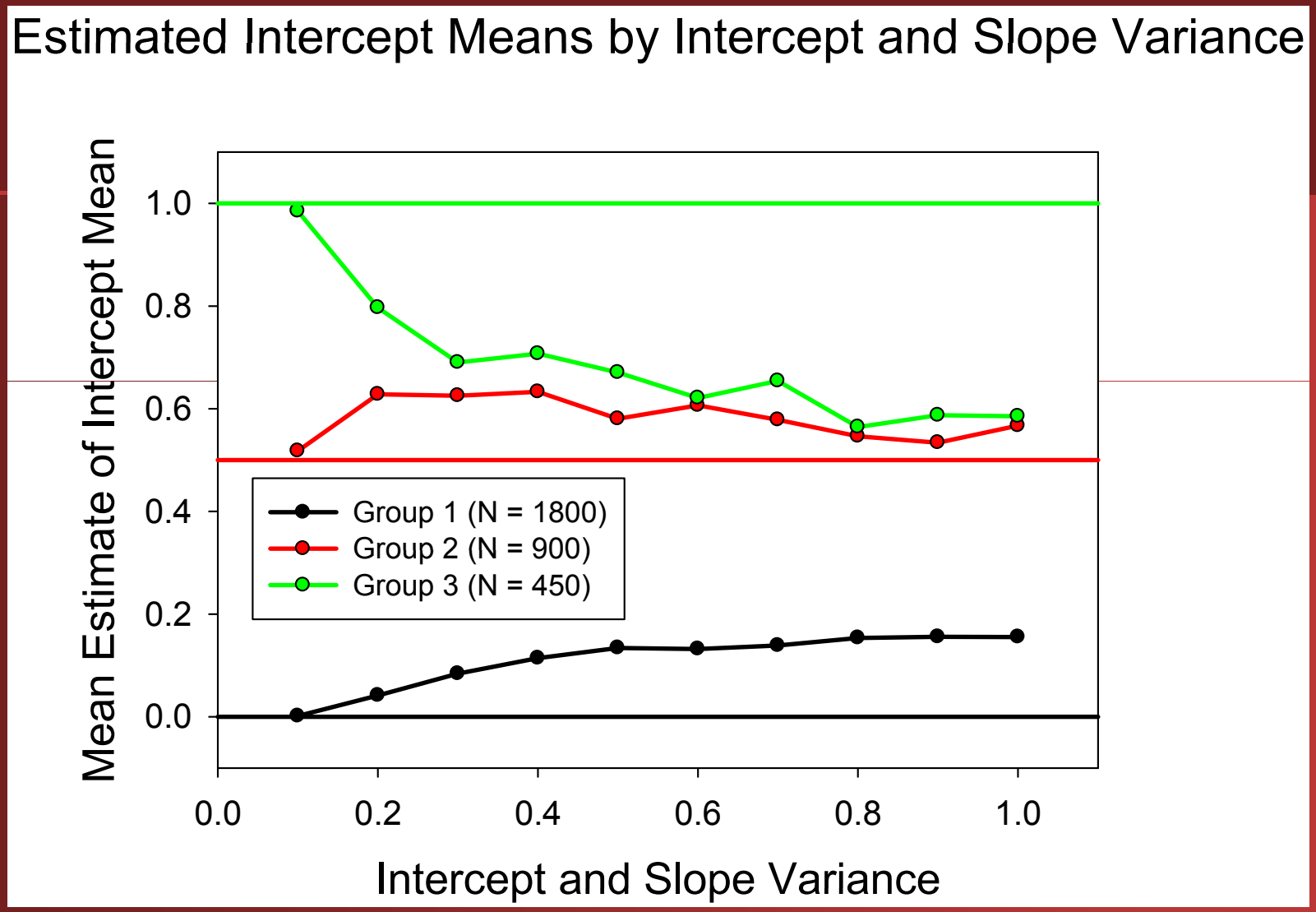
Mean Estimate of Slope Means by Intercept, Slope Effect Size

Residual variance = 0.1, with continuous covariate N(5,5) in Group 1, N(15,5) in Group 2



Conclusions

- GMMs – potentially very informative
- Currently seen mainly in drug/alcohol and developmental studies – esp. criminology
- Simulation results raise concerns
  - Poor estimation of group membership
  - Model parameters require large effect sizes


Estimated Intercept Means by Intercept and Slope Variance
Mean Estimate of Intercept Mean
1.0
0.8
0.6
0.4
0.2
0.0
Group 1 (N = 1800)
Group 2 (N = 900)
Group 3 (N = 450)
0.0
0.2
0.4
0.6
0.8
1.0
Intercept and Slope Variance