Phenomics

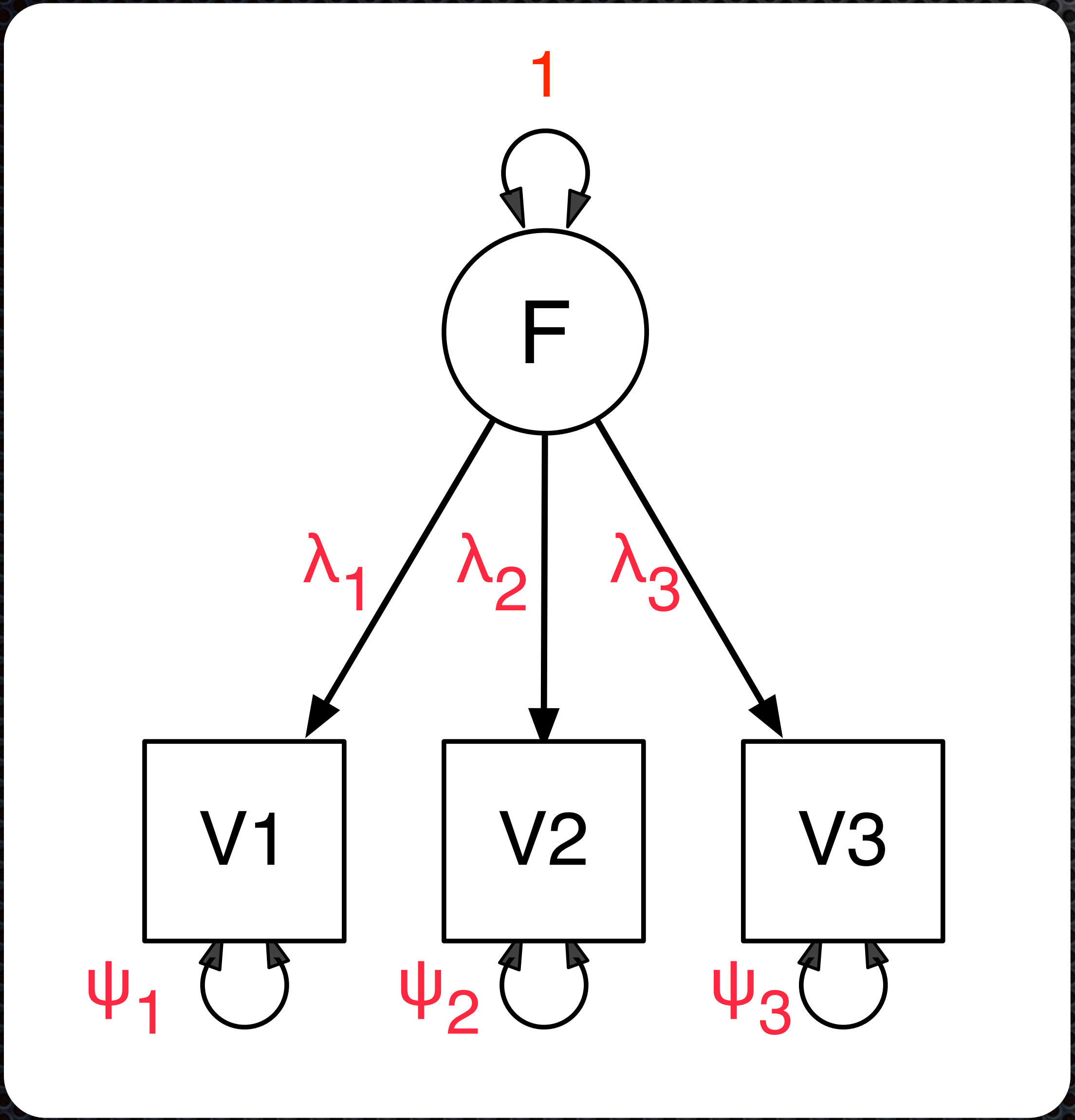
Michael C. Neale
Virginia Institute for Psychiatric & Behavioral Genetics
Virginia Commonwealth University

Boulder Workshop
10 March 2017

Outline

- Factor Analysis and Measurement Invariance
- Testing Hypotheses about Gene Action: FTND
- Drug factors vs. Symptom factors
- Genomewide Structural Equation Modeling
- Occult Subtypes
- Obligate Missingness
- Age differences decay correlations
- Future Directions including GREML

Measurement Invariance: Factor Model



Usually want to know about F, the latent factor!

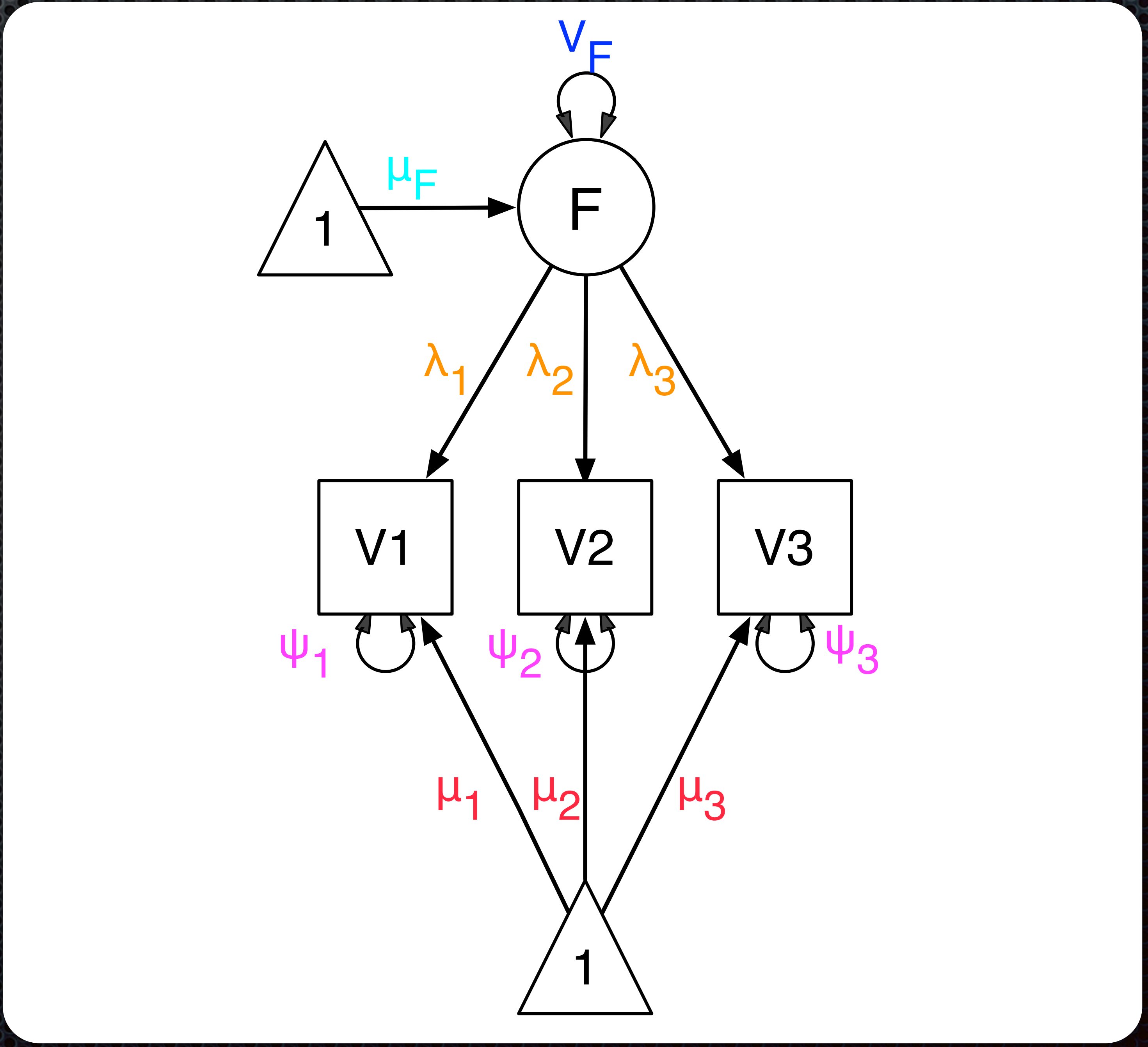
Indirect measurement

MNI Causes Errors of Inference

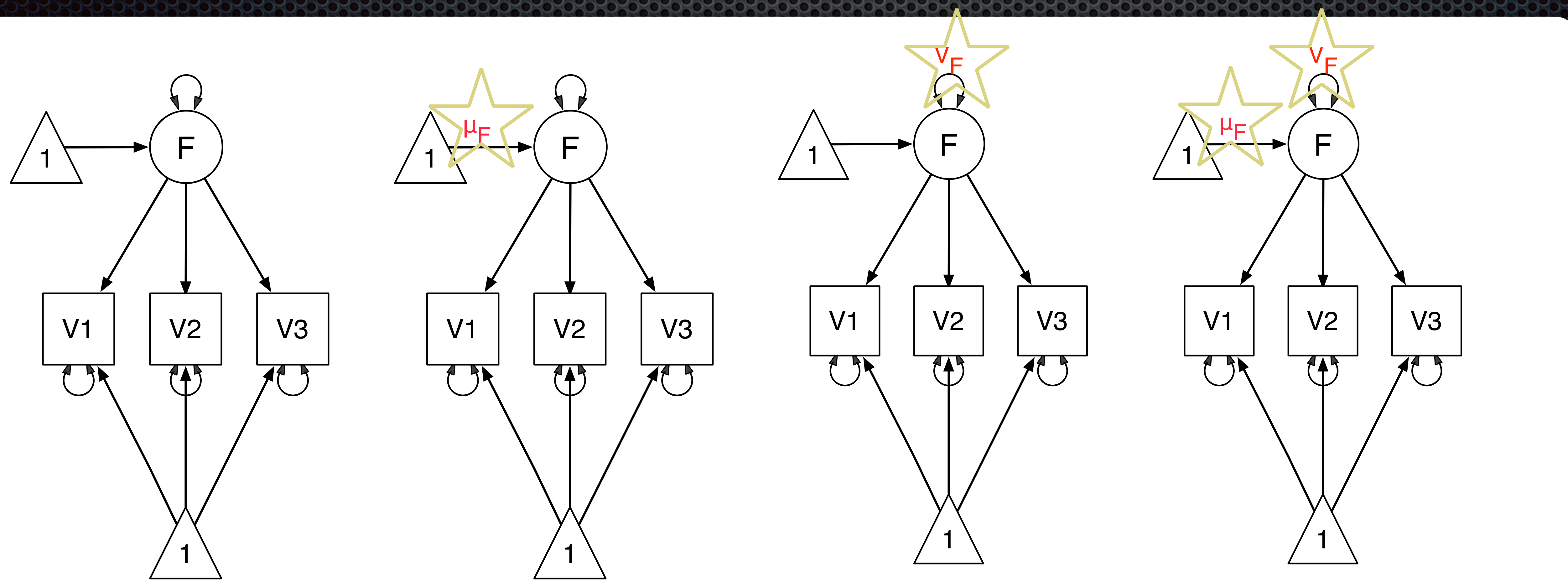
- Sum Scores & Factor Scores Depend on Model
- Item-level Differences May:
  - Invalidate Group Mean Tests (Association even)
  - Invalidate Group Variance Tests
- MI Still Rarely Tested

Invariance: Five Potential Types of Difference

- Factor Variances
- Factor Means
- Factor Loadings
- Item Variances
- Item Means



Invariance Models of Factor-Level Effects



1. No Covariates

2. Age/Sex on Factor Mean

3. Age/Sex on Factor Variance

4. Age/Sex on Factor Mean and Variance

Correlations across Substances: Add Health

	Stimulants	Tranquilizers	Marijuana
Stimulants	1		
Tranquilizers	0.74	1	
Marijuana	0.63	0.66	1
Factor Loadings	0.84	0.87	0.75

Medland & Neale (2010) An integrated phenomic approach to multivariate allelic association. European Journal of Human Genetics18:233–239

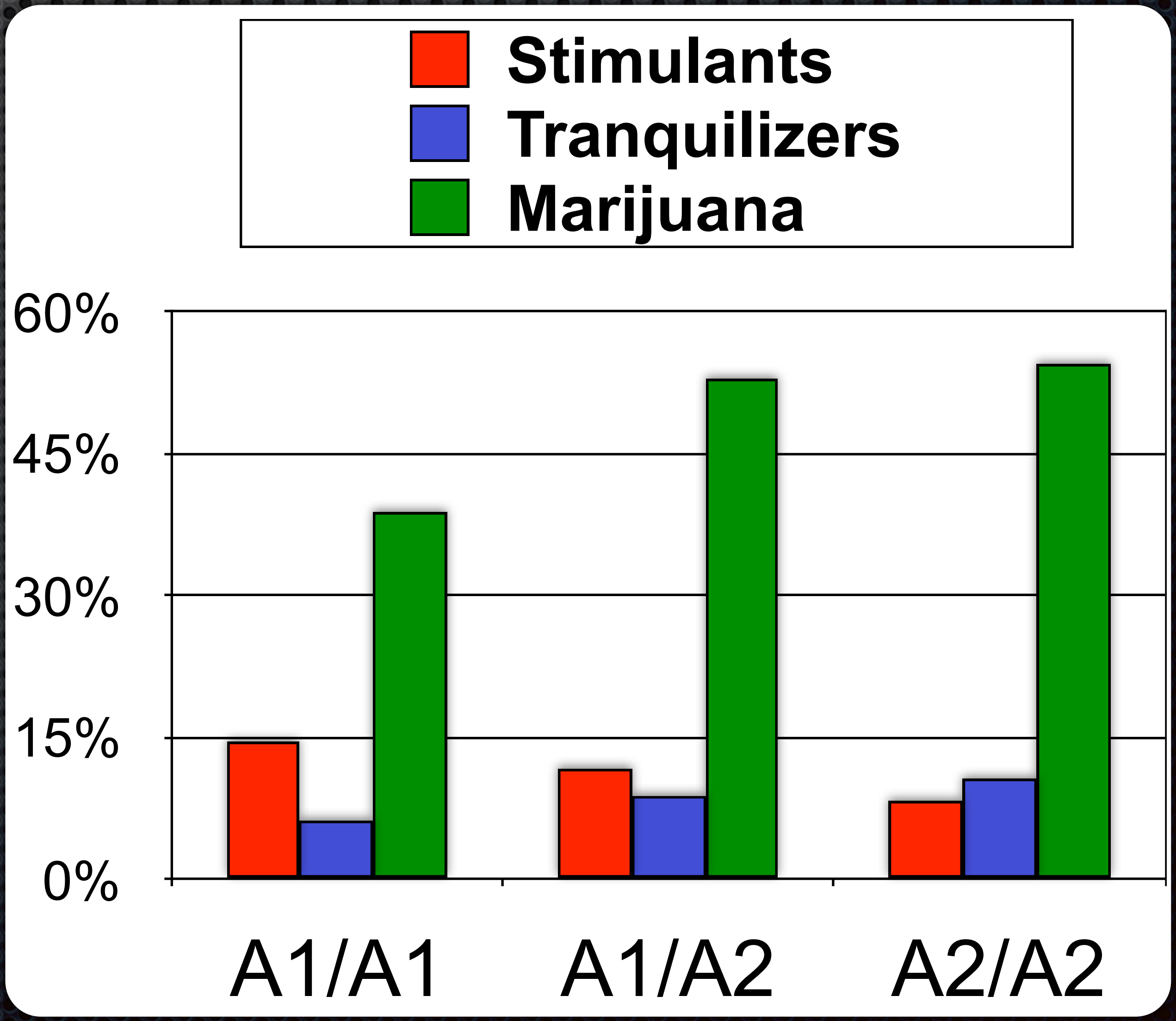
DRD2 Association Results (Add Health)

- Univariate associations
  - Stimulants: χ^2=3.88, β= -.18, p < .05
  - Tranquilizers: χ^2=1.65, β= .13, NS
  - Marijuana: χ^2=2.60, β= .11, NS
- Factor level association
  - χ^2=0.65, kF= .06, NS
- Item level association
  - χ^2=13.91 (3df; p < 0.005)

$\beta_{Stimulants}$ = -0.19

$\beta_{Tranquilizers}$ = 0.14

$\beta_{Marijuana}$ = 0.11



MI Application: National Survey of Drug Use in Households (NSDUH)

- Substance Abuse and Mental Health Services Administration (SAMSA) regular data collection
- ~50,000 persons per assessment
- Face-to-face Interviews(!)
- Audio-Computer-Assisted Testing

Map Items to DSM-IV Substance Abuse and Dependence Criteria

- A1 During the past 12 months, did using marijuana or hashish cause you to have serious problems like this either at home, work, or school?
- A2 During the past 12 months, did you regularly use marijuana or hashish and then do something where using marijuana or hashish might have put you in physical danger?
- A3 During the past 12 months, did using marijuana or hashish cause you to do things that repeatedly got you in trouble with the law?
- A4 Did you continue to use marijuana or hashish even though you thought it caused problems with family or friends?

DSM-IV Dependence Criteria

- D1 During the past 12 months, did you need to use more marijuana or hashish than you used to in order to get the effect you wanted?
- D3 Were you able to keep to the limits you set, or did you often use marijuana or hashish more than you intended to?
- D4 During the past 12 months, did you want to or try to cut down or stop using marijuana or hashish?
- D5 During the past 12 months, was there a month or more when you spent a lot of your time getting or using marijuana or hashish?
- D6 This question is about important activities such as working, going to school, taking care of children, doing fun things such as hobbies and sports, and spending time with friends and family.
  - During the past 12 months, did using marijuana or hashish cause you to give up or spend less time doing these types of important activities?
- D7 Did you continue to use marijuana or hashish even though you thought it was causing you to have physical problems?

Test of Factor Loading Invariance: Marijuana in NSDUH

Model	Npar	Comparison Model	Likelihood	df	AIC	diffLL	diffdf	p	Age Effect	Sex Effect
1. No Covariates	20	NA	62514	78204	-93894	NA	NA	NA	NA	NA
2. Age/Sex on Factor Mean	22	1	62009	78202	-94395	505	2	<.0001	-3.85567	-0.09112
3. Age/Sex on Factor Variance	22	1	62480	78202	-93924	33.84	2	<.0001	0.70624	0.40302
4. Age/Sex on Factor Mean and Variance	24	1	61893	78200	-94507	620.4	4	<.0001		
5. Age/Sex on Mean and Loadings	42	4	61801	78182	-94563	92.34	18	<.0001		
6. Age/Sex on Thresholds and Factor Variance	42	4	61802	78182	-94562	91.2	18	<.0001		

Test of Item Mean Invariance: Marijuana in NSDUH

Model	Npar	Comparison Model	Likelihood	df	AIC	diffLL	diffdf	p	Age Effect	Sex Effect
1. No Covariates (1f model)	20	NA	62514	78204	-93894	NA	NA	NA	NA	NA
2. Age/Sex on Factor Mean	22	1	62009	78202	-94395	505	2	<.0001	-3.85567	-0.09112
3. Age/Sex on Factor Variance	22	1	62480	78202	-93924	33.84	2	<.0001	0.70624	0.40302
4. Age/Sex on Factor Mean and Variance	24	1	61893	78200	-94507	620.4	4	<.0001		
5. Age/Sex on Mean and Loadings	42	4	61801	78182	-94563	92.34	18	<.0001		
6. Age/Sex on Thresholds and Factor Variance	42	4	61802	78182	-94562	91.2	18	<.0001		

Test of Item Mean Invariance: Marijuana in NSDUH

- Strong evidence of MNI with respect to age and sex
- Examine individual items
- Four column heatmap for significance of effects
  - Item Means & Factor Variances
  - Sex and Age
- Compare across self-reported race

-2lnL Likelihood Ratio Test Statistics: Marijuana Item Means & Factor Loadings

Entire Sample

	Sex: IM_sex	Sex: FL_sex	Age: IM_age	Age: FL_age	Item	
	10.3	4.68	−14.28	−5.33	MRJA1	Work
	0.04	−0.05	−1.82	1.81	MRJA2	Danger
	−20.7	0	−5.94	−2.28	MRJA3	Law
	1	0	−0.41	−2.13	MRJA4	Friends
	0.12	−1.32	−11.24	−11.96	MRJD1	Tol
	−0.72	−1.53	14.87	−2.74	MRJD3	>Intend
	0.69	−0.37	28.87	12.54	MRJD4	TryCut
	−11.52	6.18	0	−0.09	MRJD5	TimeGet
	1.8	−0.09	14.79	−14.2	MRJD6	TimeOther<
	0.81	−1.94	16.17	−9.31	MRJD7	PhysProb

+/- sign denotes direction

-2lnL Likelihood Ratio Test Statistics: Marijuana Item Means & Factor Loadings

White

	Sex		Age			
	IM_sex	FL_sex	IM_age	FL_age		
	3	3.46	−3.85	−2.66	MRJA1	Work
	1.69	−0.11	−6.87	1.34	MRJA2	Danger
	−7.52	−0.12	−5.82	−2.2	MRJA3	Law
	1.68	1.8	−1.21	−1.04	MRJA4	Friends
	0.09	−1.93	−10.72	10.29	MRJD1	Tol
	−0.05	−0.23	14.99	−6.57	MRJD3	>Intend
	2.81	−0.1	23.1	−15.97	MRJD4	TryCut
	−6.33	1.12	−0.25	0	MRJD5	TimeGet
	−0.04	0.9	17.45	−12.71	MRJD6	TimeOther<
	0.02	−1.68	12.77	−12.62	MRJD7	PhysProb

-2lnL Likelihood Ratio Test Statistics: Marijuana Item Means & Factor Loadings

African American

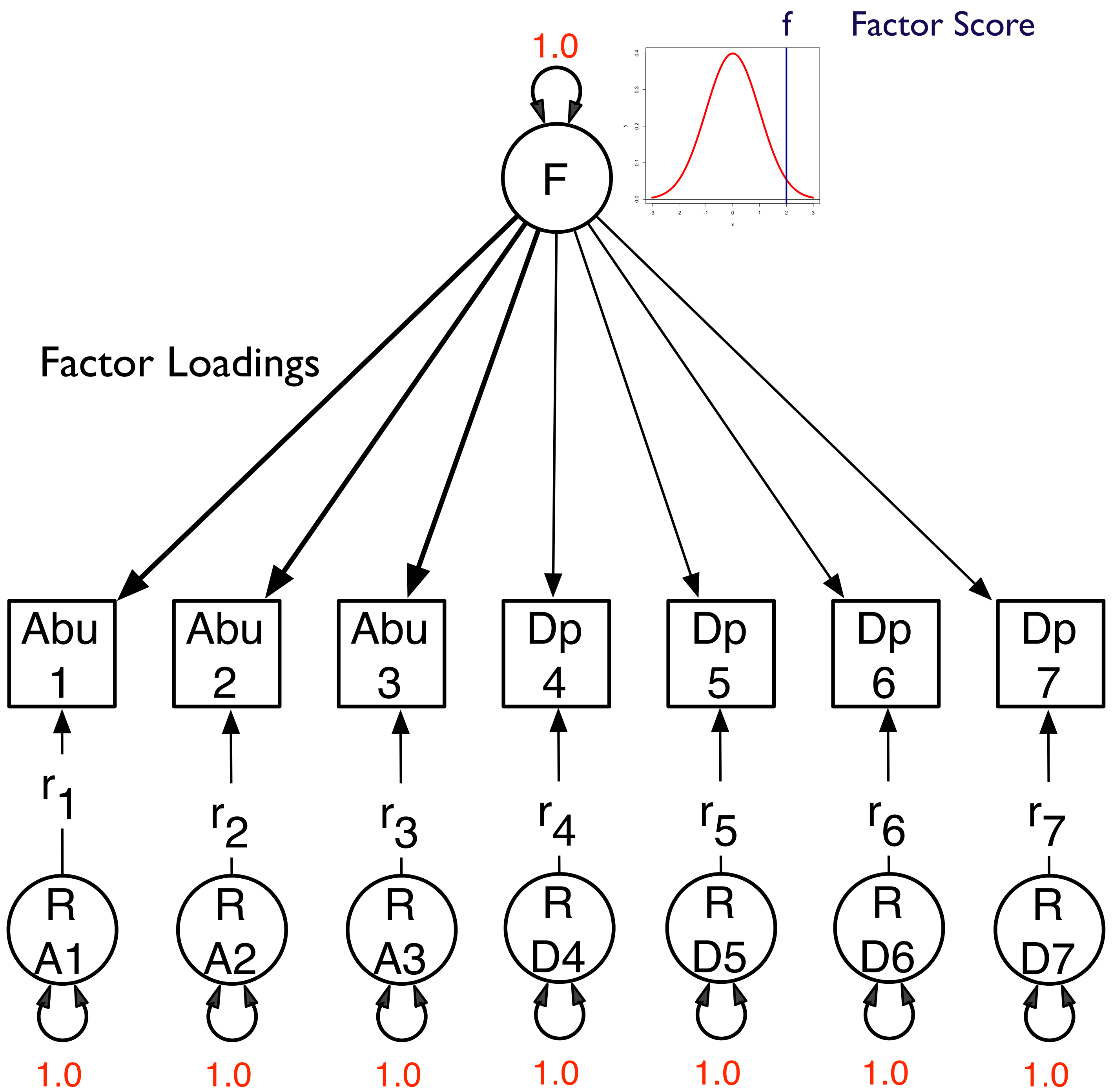
	Sex		Age			
	IM_sex	FL_sex	IM_age	FL_age		
	2.52	6.42	−1.67	−2.87	MRJA1	Work
	−0.04	−0.26	0.31	−0.21	MRJA2	Danger
	−15.34	−11.7	0.28	0.18	MRJA3	Law
	−1.07	−2.63	0.58	−0.13	MRJA4	Friends
	0.58	−1.23	−0.15	0.04	MRJD1	Tol
	0	−2.01	−0.06	0.84	MRJD3	>Intend
	−0.33	−0.64	1.35	0	MRJD4	TryCut
	−5.1	4.49	−0.01	0.01	MRJD5	TimeGet
	10.34	6.78	−0.07	0.42	MRJD6	TimeOther<
	0.11	0	0.17	−0.37	MRJD7	PhysProb

-2lnL Likelihood Ratio Test Statistics: Marijuana Item Means & Factor Loadings

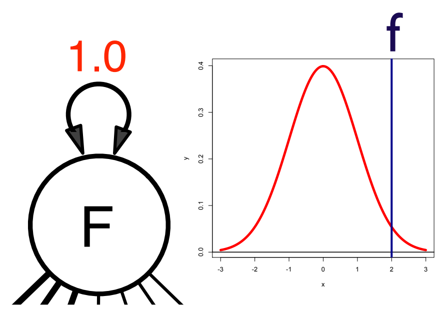
Hispanic

	Sex		Age			
	IM_sex	FL_sex	IM_age	FL_age		
	8.9	−1.36	−12.16	4.74	MRJA1	Work
	−0.52	−0.06	0.01	2.43	MRJA2	Danger
	−11.54	1.87	−1.36	−0.29	MRJA3	Law
	0.37	−2.95	−0.31	−0.12	MRJA4	Friends
	−0.28	0.81	−0.91	1.92	MRJD1	Tol
	−1.56	0.19	2.77	0	MRJD3	>Intend
	0.13	−0.22	3.32	0.24	MRJD4	TryCut
	−1.14	1.26	0.57	−0.19	MRJD5	TimeGet
	0.5	−1.26	3.49	−4.03	MRJD6	TimeOther<
	1.27	−0.07	4.3	−0.1	MRJD7	PhysProb

Estimating Factor Scores



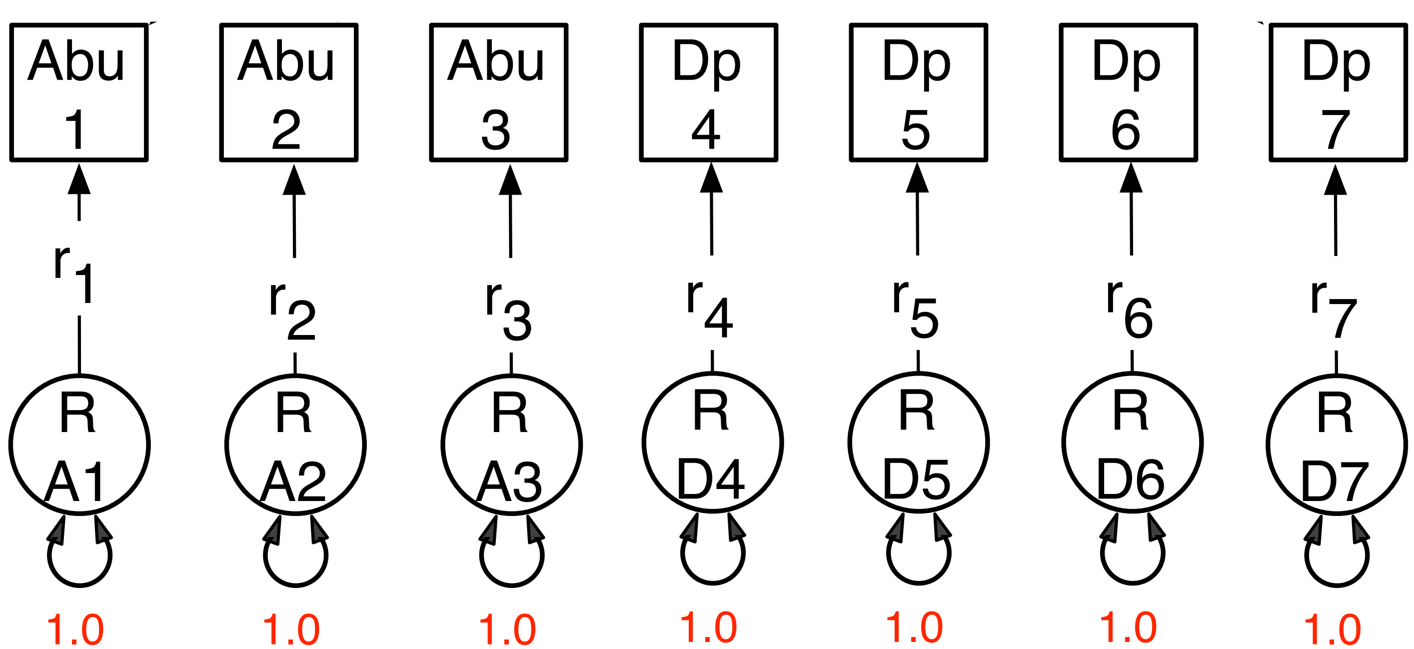
ML Estimation of Factor Scores

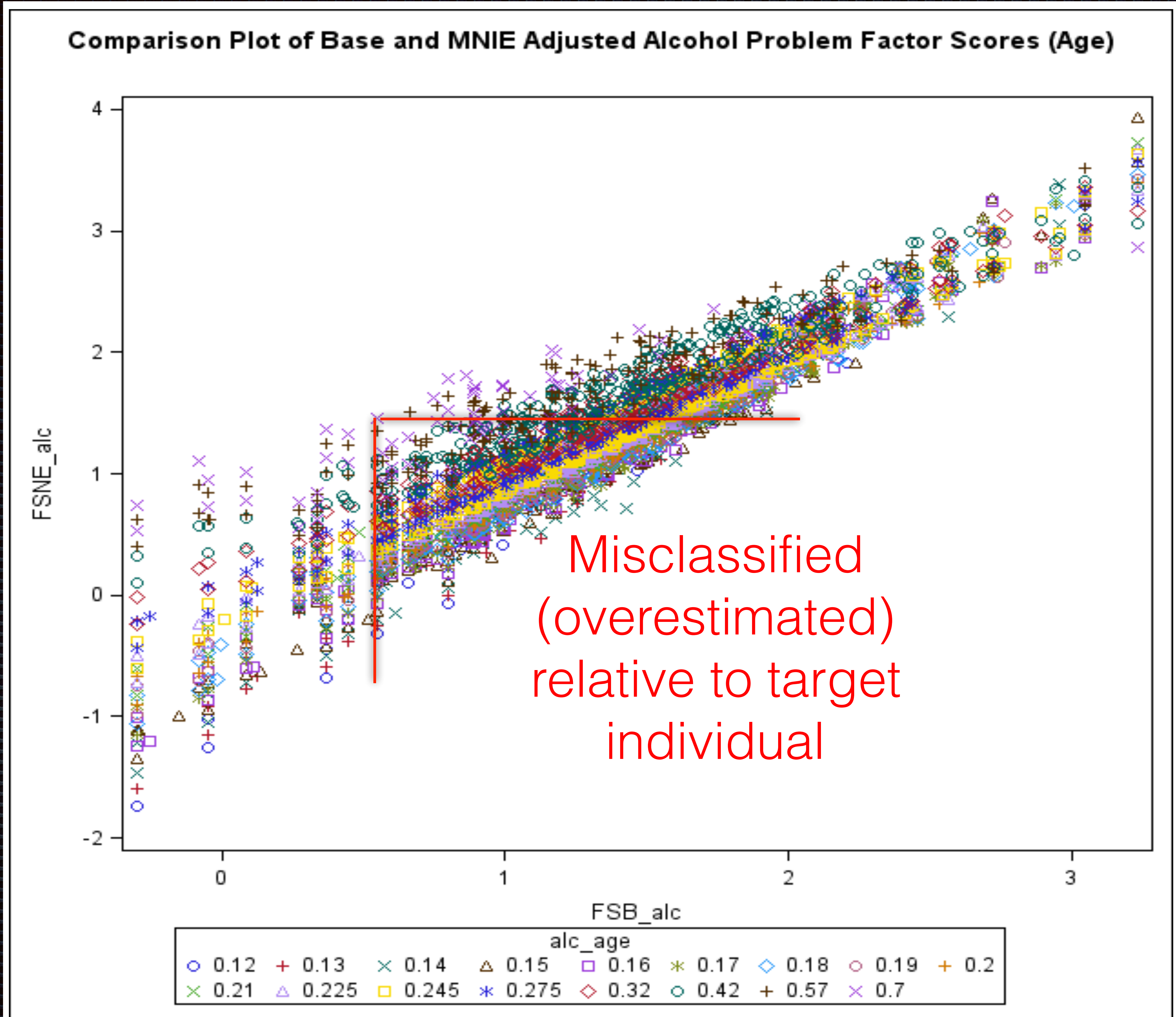


Factor Score

Factor Score * Likelihood of items conditional on factor score

Items independent conditional on factor score:
Means and variances change according to size of factor loadings



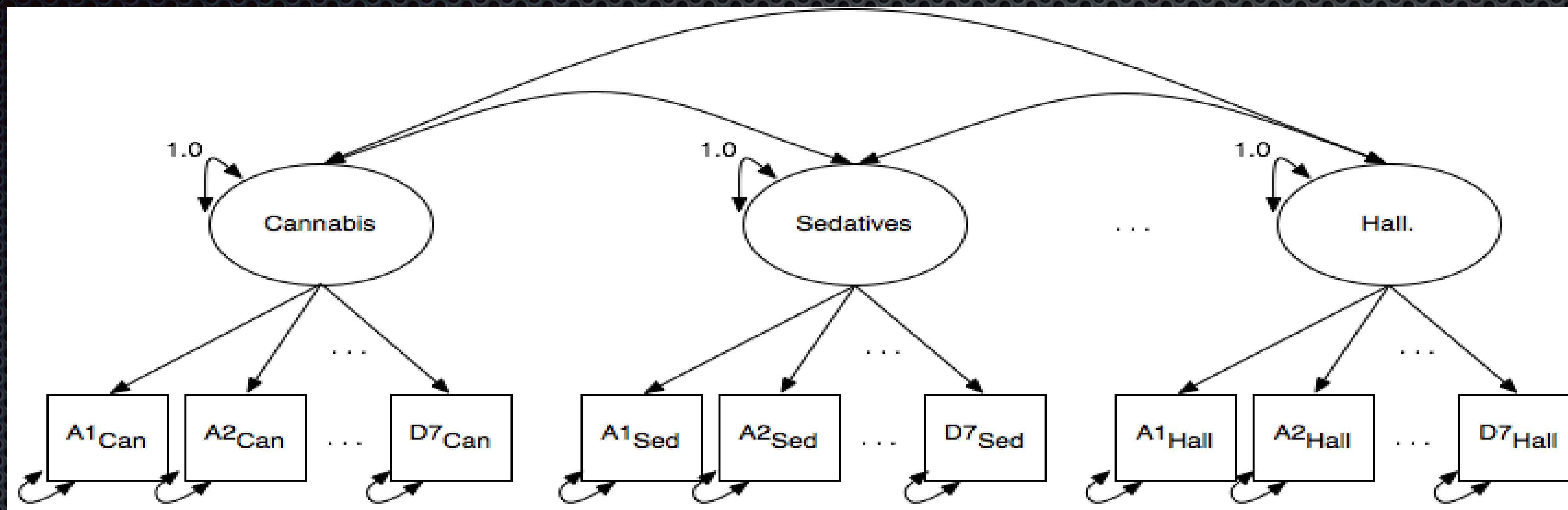


Comparison Plot of Base and MNIE Adjusted Alcohol Problem Factor Scores (Age)
FSNE_alc
4
3
2
1
0
-1
-2
FSB_alc
0
1
2
3
alc_age
0.12
0.13
0.14
0.15
0.16
0.17
0.18
0.19
0.2
0.21
0.225
0.245
0.275
0.32
0.42
0.57
0.7
Misclassified (overestimated) relative to target individual

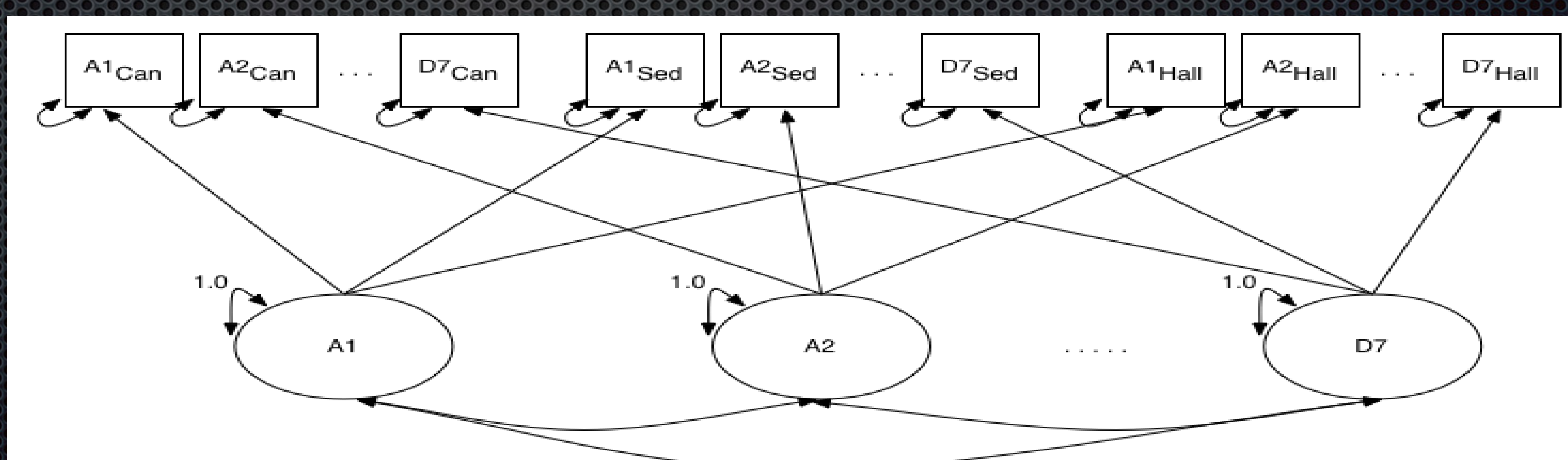
Drug vs Symptom Factors

- DSM III-R/IV drug abuse and dependence symptoms for cannabis, sedatives, stimulants, cocaine, opioids and hallucinogens
- 13 misuse symptoms measured across six illicit substance categories (78 items)
- 4179 males born 1940–1970 from the population-based Virginia Adult Twin Study of Psychiatric and Substance Use Disorders
- Confirmatory factor analyses tested specific hypotheses regarding the latent structure of substance misuse

Drug vs Symptom Factors



Clark, S. L., Gillespie, N. A., Adkins, D. E., Kendler, K. S., and Neale, M. C. (2016). Psychometric modeling of abuse and dependence symptoms across six illicit substances indicates novel dimensions of misuse. *Addict Behav*, 53:132–40. PMCID: PMC4679450.



Drug vs Symptom Factors

Model	χ^2	DF	*p*-Value	CFI	RMSEA
M1: Drug factors only	4175	2910	<0.001	0.78	0.017
M2: Misuse characteristic factors only	3647	2847	<0.001	0.86	0.013
M3: Drug and misuse characteristic factors	2966	2754	<0.001	0.96	0.007
M4: General liability factor	4598	2925	<0.001	0.71	0.019
M1 vs. M3	1209	156	<0.001		
M2 vs. M3	681	93	<0.001		

- Adding symptom factors dramatically improves fit
- Majority of variance in many Sx due to symptom not drug factor

Factor Score Notes

- Factor scores do not all have same error variance
- Factor scores of A, C & E components may correlate highly
- Latent trait may be non-normal (Schmitt et al 2006 Multiv Behav Res)
- Factor loadings (precision) may vary across the distribution and give spurious GxE results
- Variation may be discrete not continuous

Item Response Probability

Example item response probability shown in white

Possible population distribution in green

AFQT

100 Items

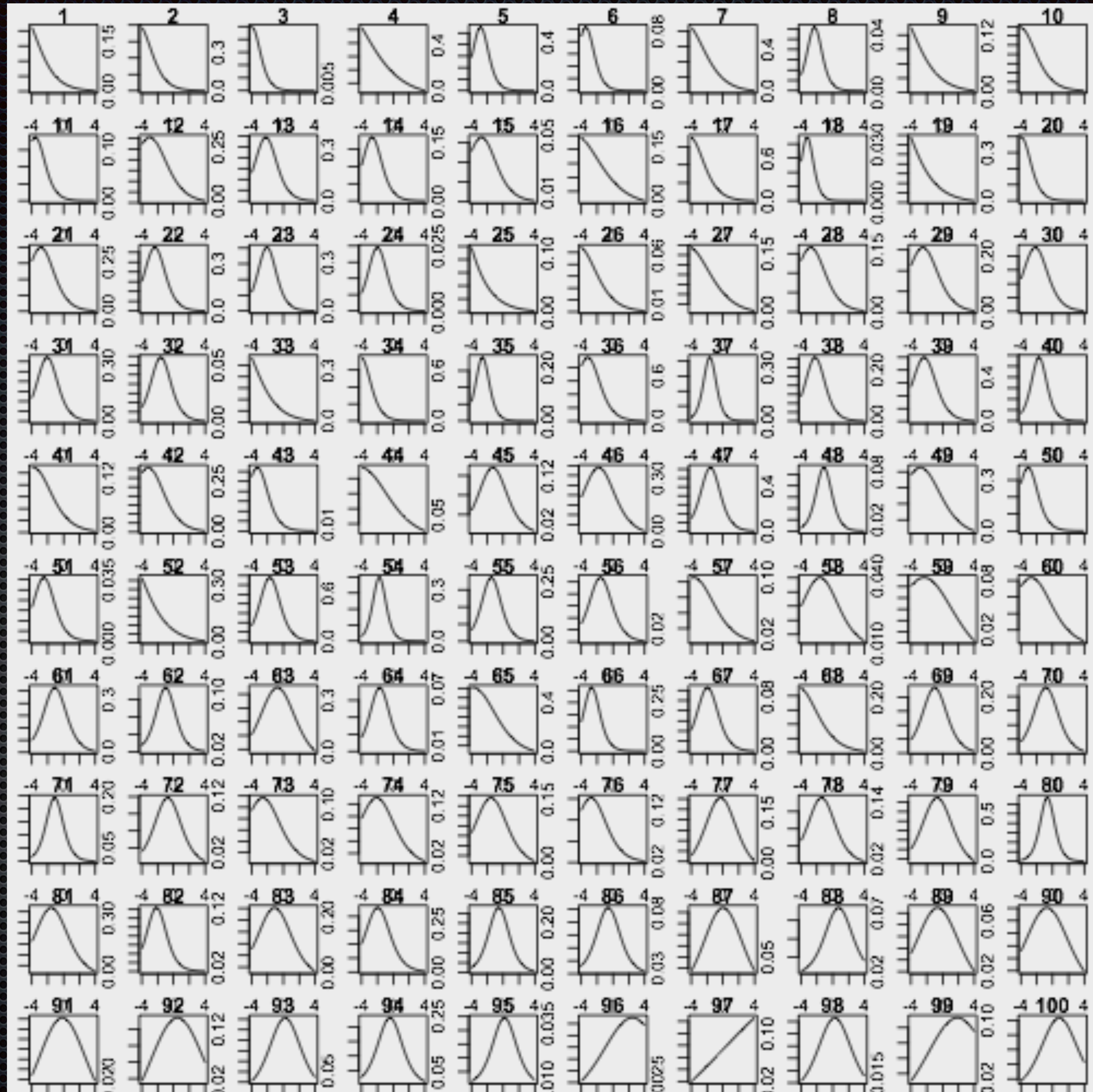
Subscales

1 Arithmetic Reasoning

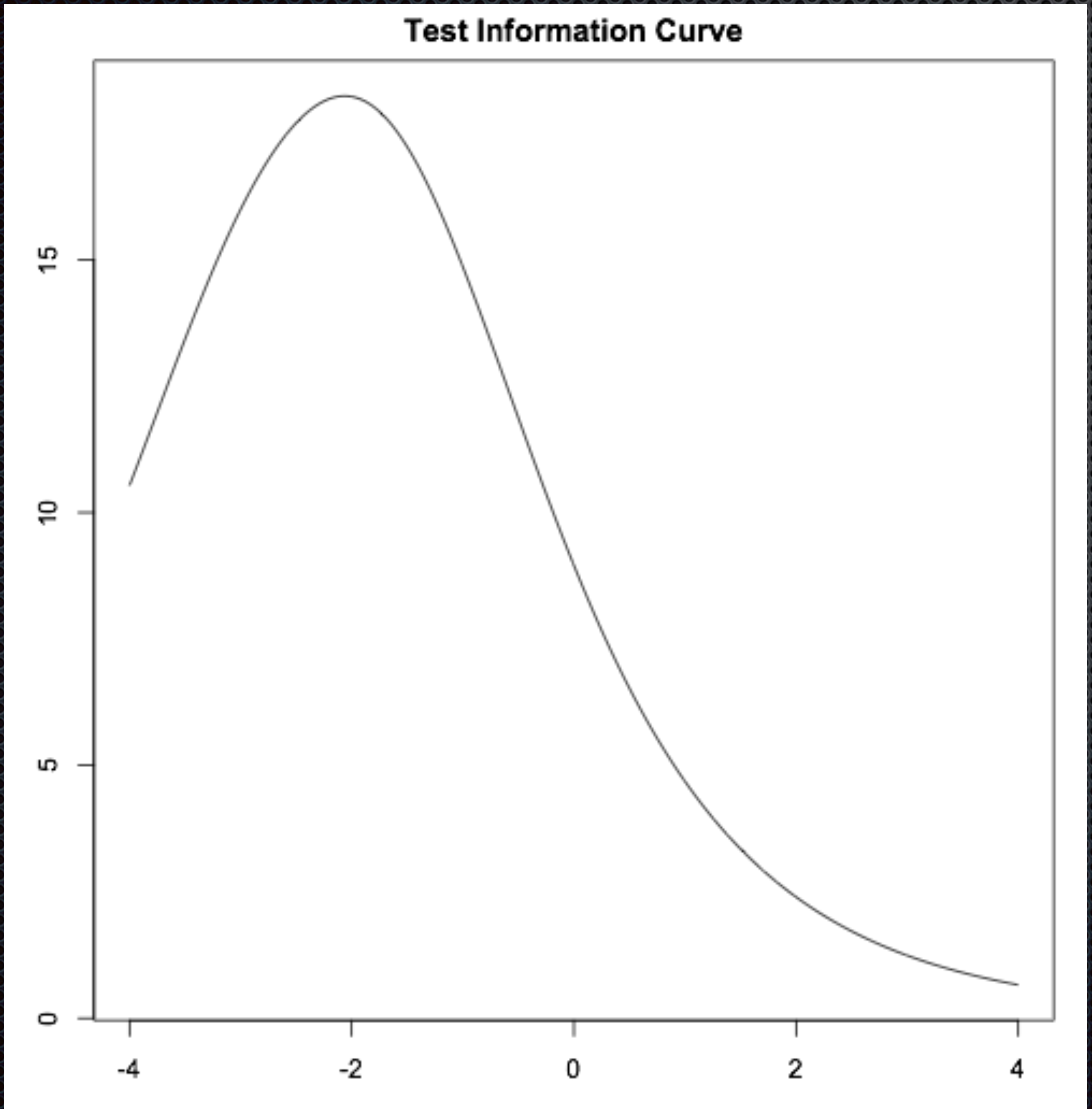
2 Mathematics Knowledge

3 Word Knowledge

4 Paragraph Comprehens.



AFQT: Overall Test Information Curve



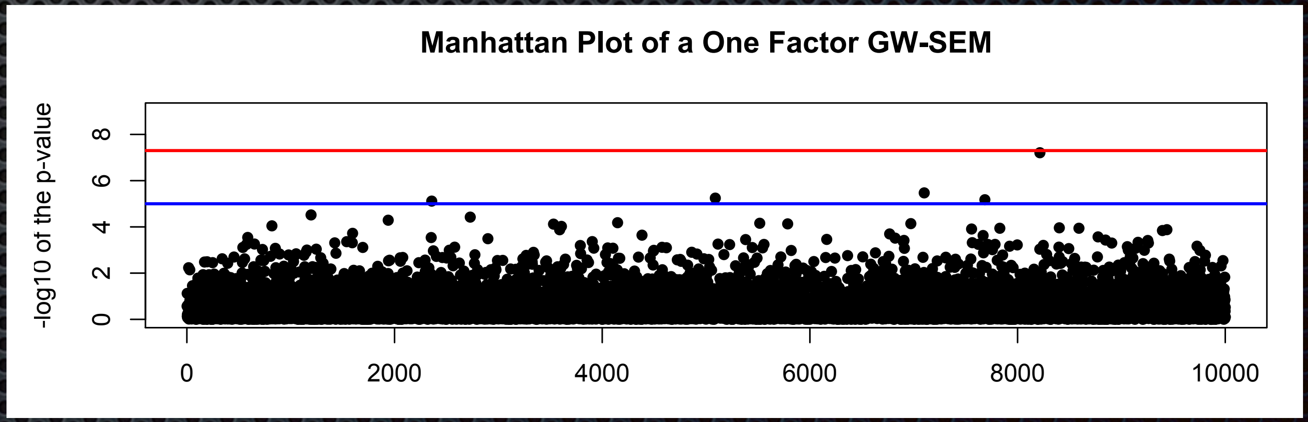
More information at left

By design

Consequences for GxE?

Genome-wide SEM

- Avoid problems with factor scores
- Fit factor or growth curve models to ordinal data
- Include effect of SNP on factor or items
- Repeat for the other 8m-1 SNPs
- Manhattan plot results
- http://goo.gl/f44UmD



Verhulst, B, Maes, H, & Neale, M (In Press) GW-SEM: A Statistical Package to Conduct Genome-Wide Structural Equation Modeling. *Behavior Genetics*

Testing Hypotheses about Gene Action: FTND

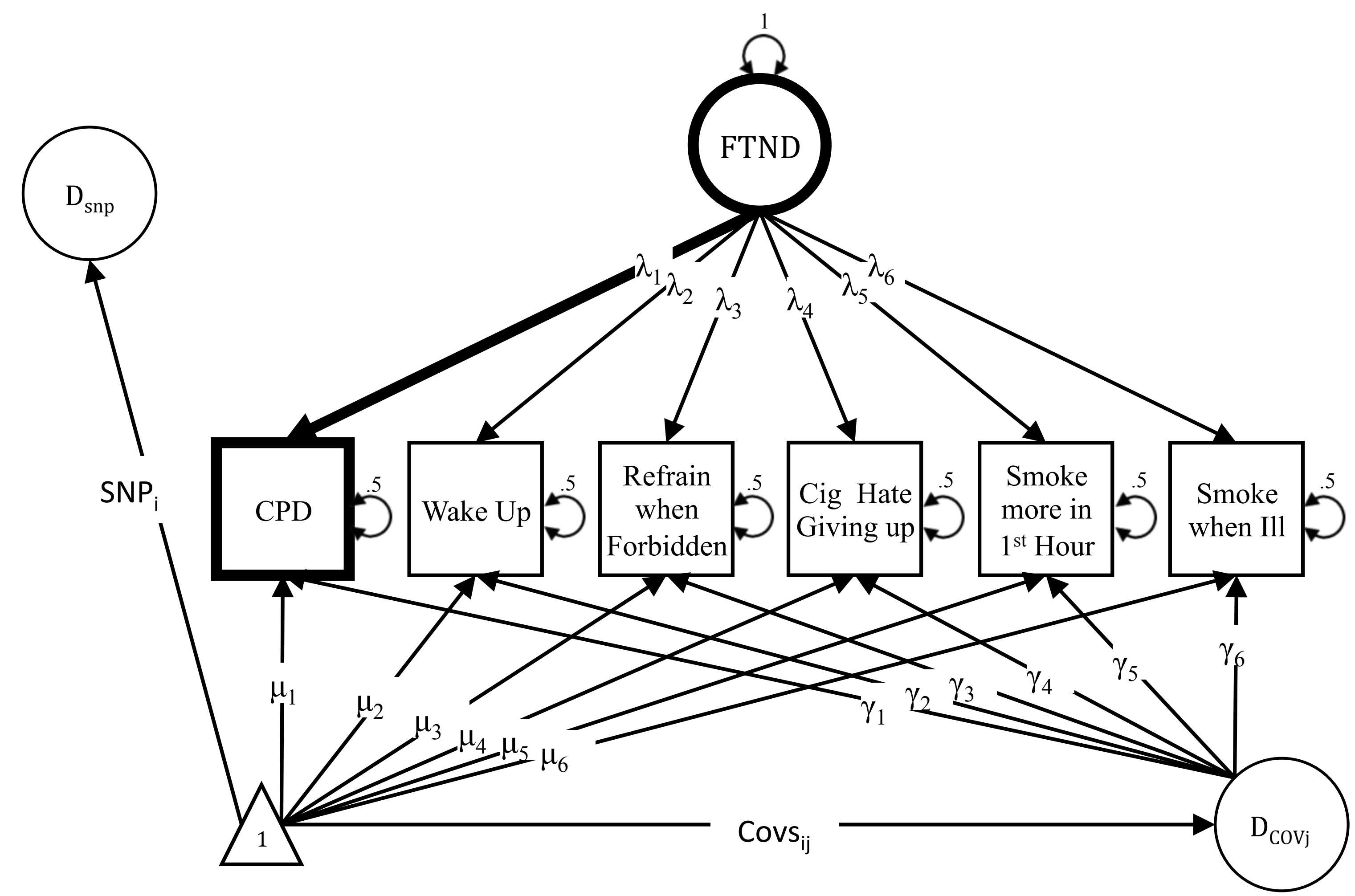
Table 1: Percentage of Variance Accounted for by the SNP rs16969968 in Latent FTND and Measured CPD

Sample	N	FTND	Total CPD	Indirect Effect
Sage	2,461	0.46	1.70	0.08
Smoking Cessation (SC)	574	0.48	1.76	0.08
CIDR	296	0.50	1.85	0.08
COPD	2,042	0.45	1.67	0.08

Note: The Direct Effects of FTND and the Total and Indirect Effects on CPD are taken from the best fitting model (H_{1c}).

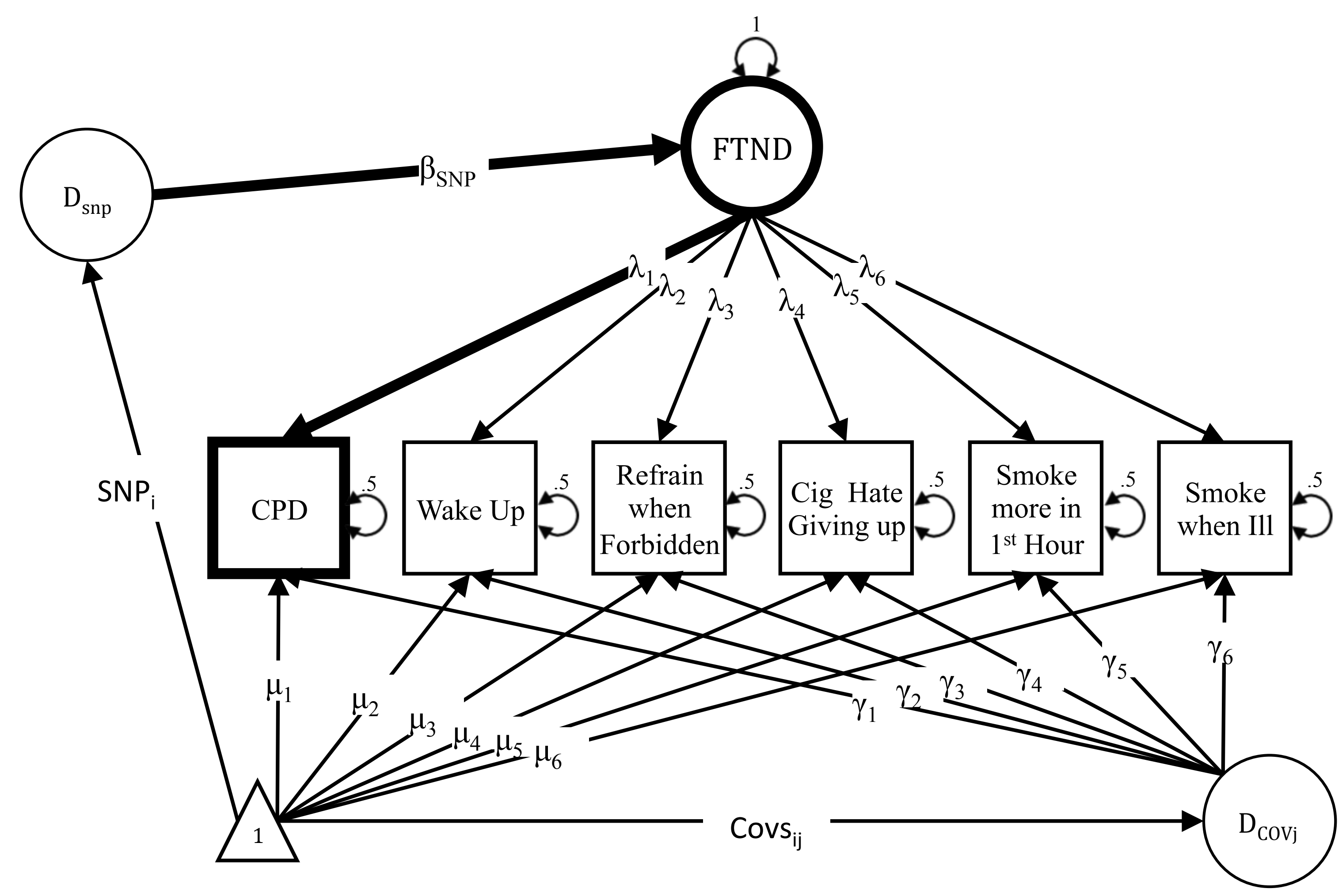
- rs16969968 Neuronal acetylcholine receptor subunit α-5 CHRNA5 associated with both ND and CPD
- What is the mechanism of action?
- CPD mere symptom of FTND
- Increases CPD increases addiction?
- Feedback loop between CPD and addiction?

H_0 No Association



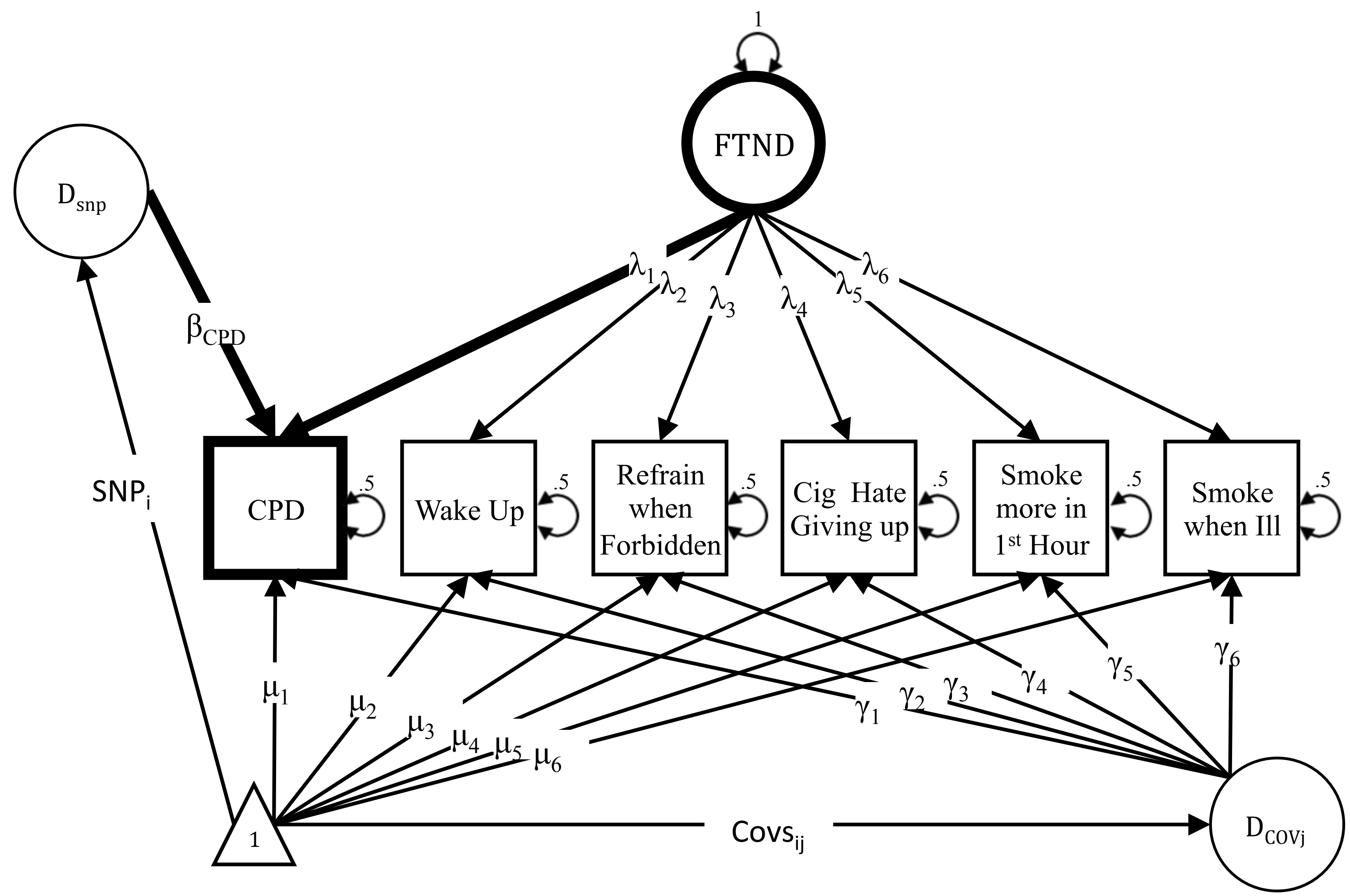
(a) H_0: Path diagram showing the Null Model where rs16969968 is unrelated to latent FTND factor and CPD.

H_{1a} Factor Only



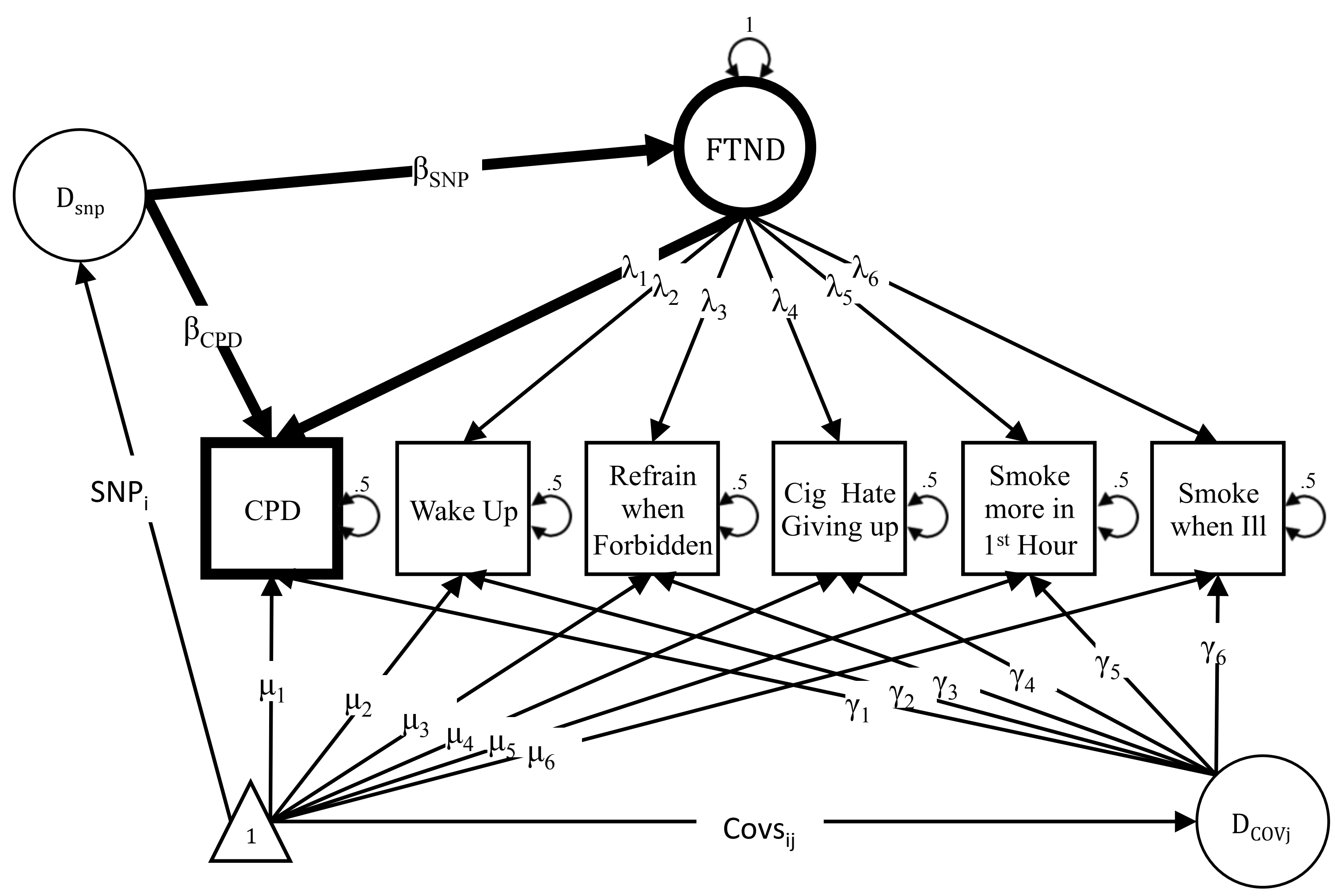
(b) H_{1a}: Path diagram for regression of the latent FTND factor on the SNP.

H_{1b} CPD Only



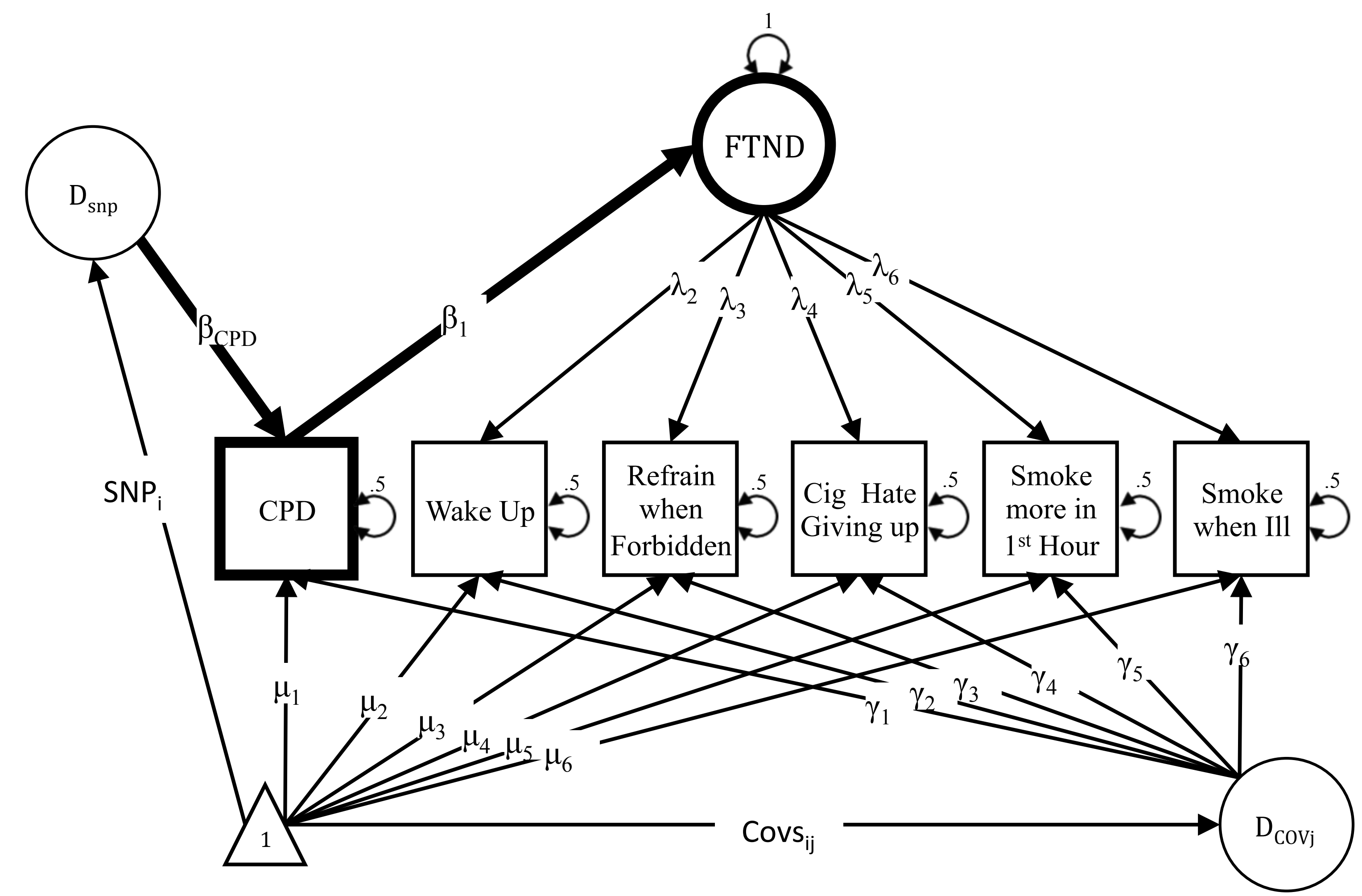
(c) H_{1b}: Path diagram for regression of CPD on the SNP

H_{1c} Factor & CPD



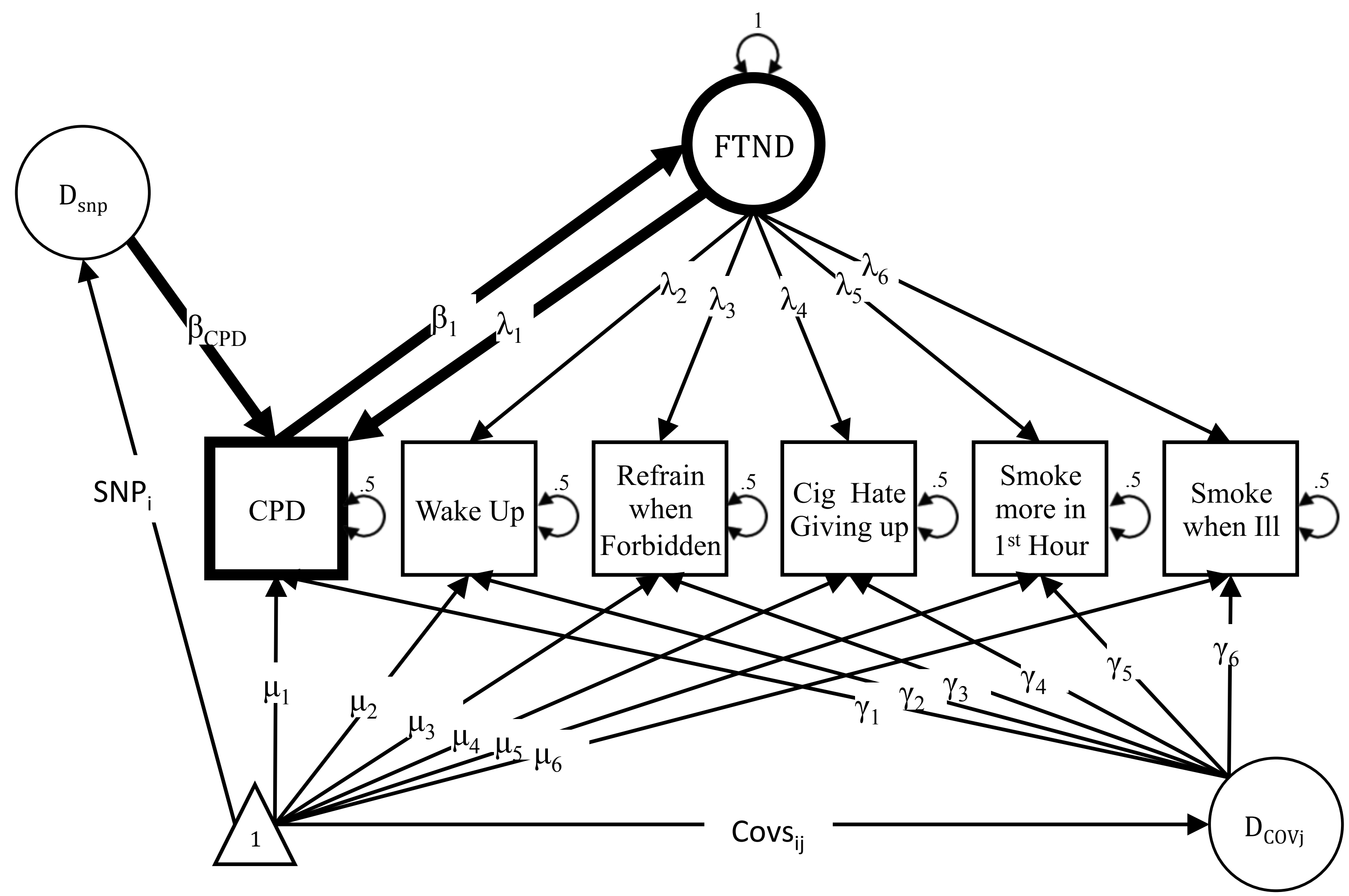
(d) H_{1c}: Path diagram for regression of the latent FTND factor and CPD on the SNP

H_{2a} CPD Only & CPD causes Factor



(e) H_{2a}: Path diagram for sequential effect of the SNP causing CPD, which causes Nicotine Dependence.

H_{2b} SNP to CPD & Reciprocal Factor



(f) H_{2b}: Path diagram for the the SNP causing CPD, which reciprocally causes Nicotine Dependence.

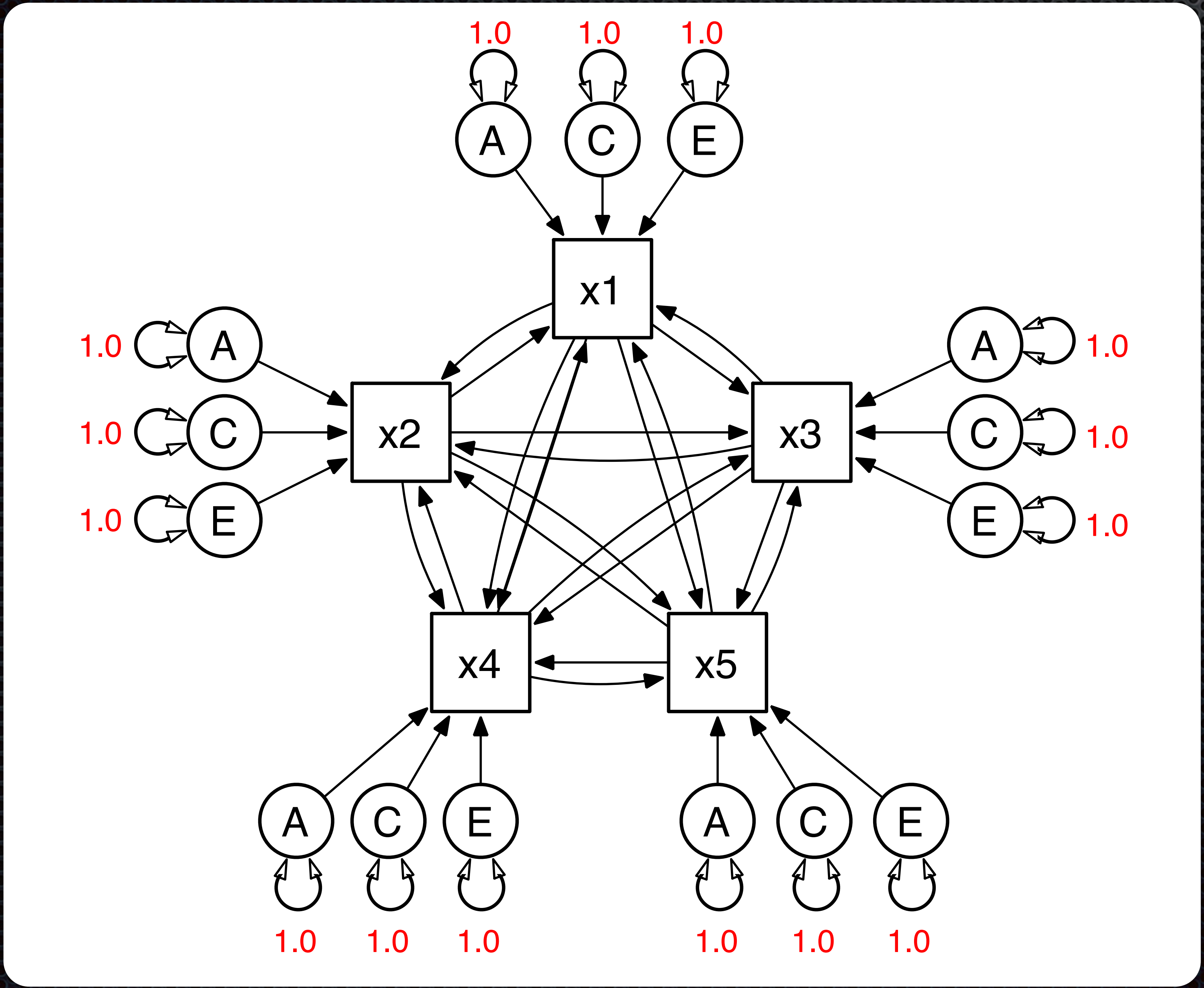
Results

Hypothesis	SNP Effect	ep	-2LL	AIC	ΔLL	Δdf	p
H_{1c}	CPD & Factor	270	45217.89	-18718.11			
H_{1a}	Factor Only	269	45225.07	-18712.93	7.18	1	7.38e-03†
H_{1b}	CPD Only	269	45238.02	-18699.98	20.13	1	7.25e-06†
H_{2b}	Reciprocal	270	45238.01	-18697.99			
H_{2a}	Sequential	269	45238.02	-18699.98	0.01	1	0.93‡
H_0	No SNP Effect	268	45251.58	-18688.42	33.69	2	4.84e-08†
					13.57	2	1.13e-03‡

- No support for sequential or reciprocal hypotheses
- Rapid habit development & later assessment may obscure relationship of CPD to addiction

Verhulst, B, Neale, M, Chen, J & Chen, S (In Revision) Disentangling genetic influences on Nicotine Dependence and Quantity of Use

Factor Model Alternative: Mutualism



Identified with data from relatives

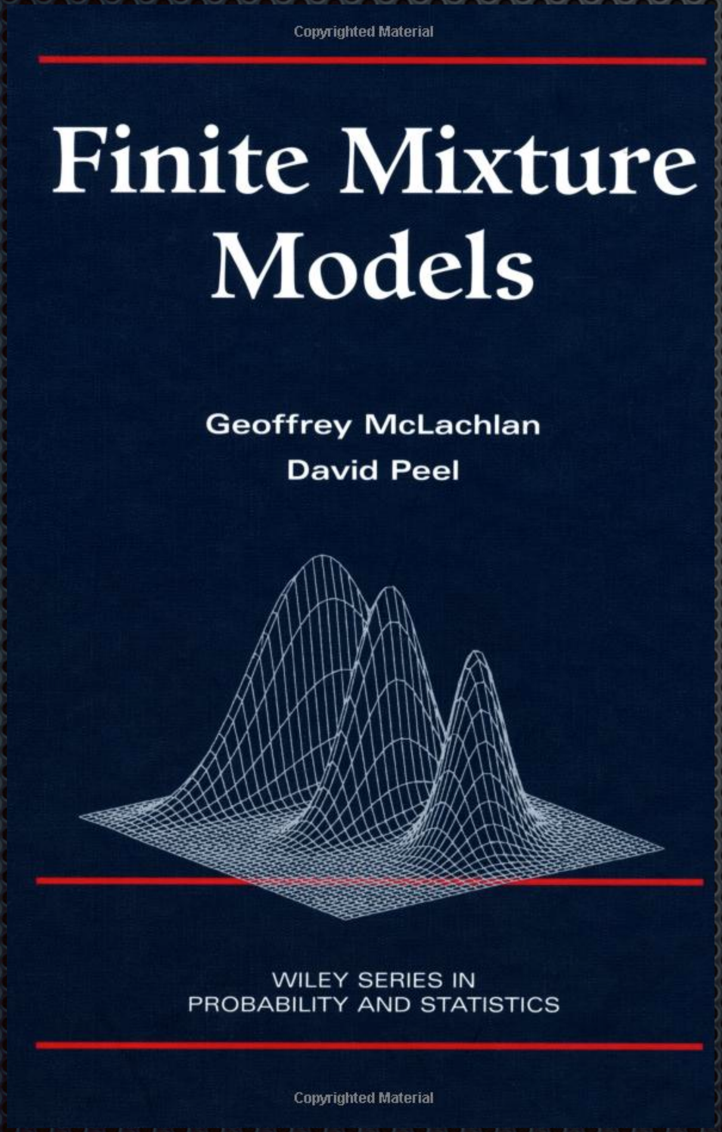
MZ & DZ Twins or adoptees needed for A/C resolution

What if Variation is Discrete?

- Latent Class and Latent Profile Models
- Factor Mixture Models
- Latent Growth Curve Mixture Models
- Regime Switching

Mixture Distributions

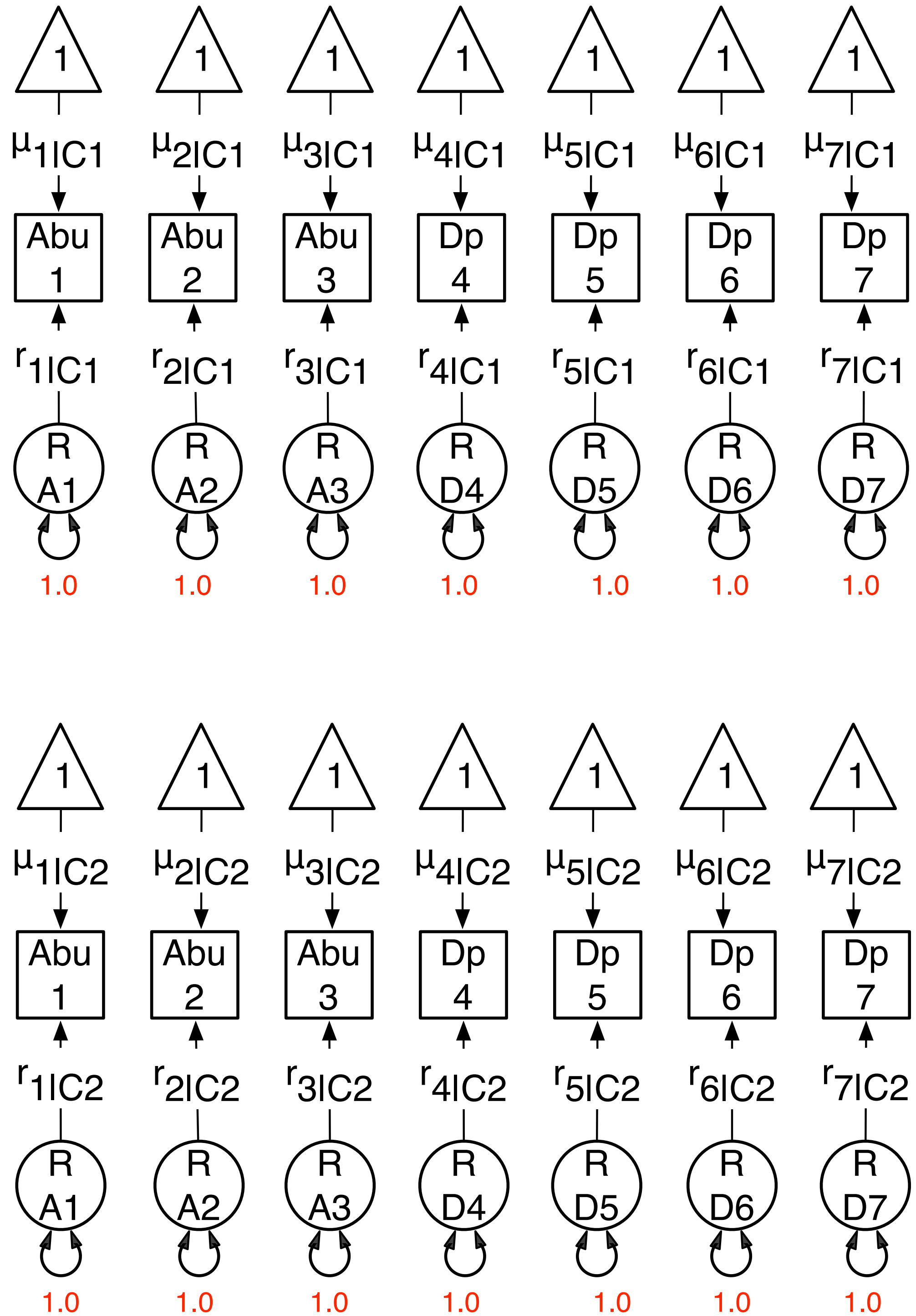
Pearson, K. (1894). Contributions to the mathematical theory of evolution. II. skew variation in homogeneous material. Philosophical Transactions of the Royal Society of London A, 186, 343-414.



- Skewness in a set of measurements of the ratio of forehead to body length of crabs
- Two species or one?

Latent Class (Subgroup)

Class 1
probability
p



Class 2
probability
(1-*p*)

Conditionally Independent?!

Expensive!

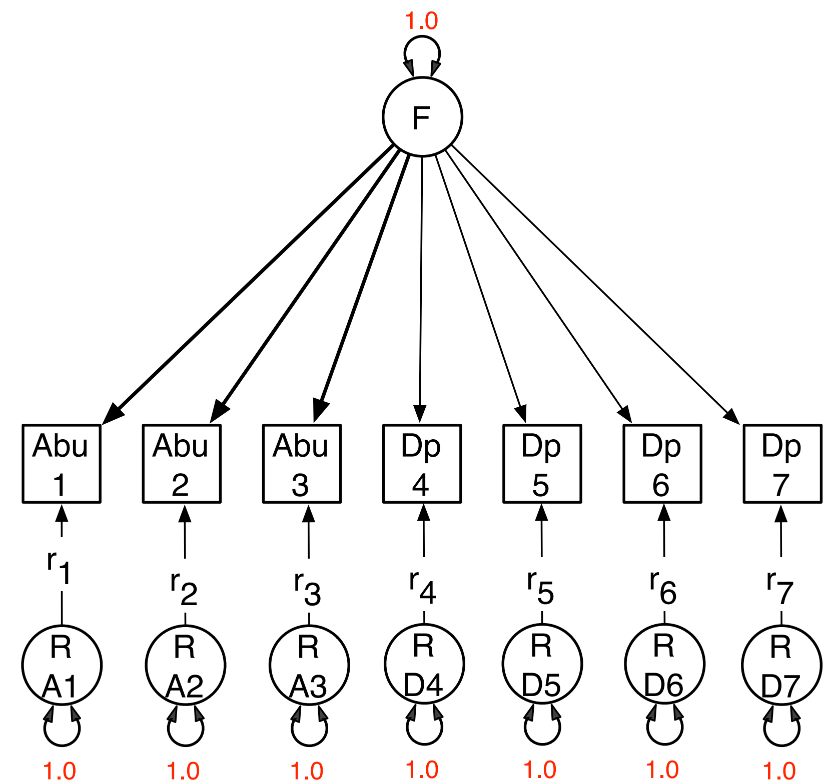
Published in final edited form as:
Int J Methods Psychiatr Res. 2010 June ; 19(2): 63–73. doi:10.1002/mpr.301.

Searching For Valid Psychiatric Phenotypes: Discrete Latent Variable Models

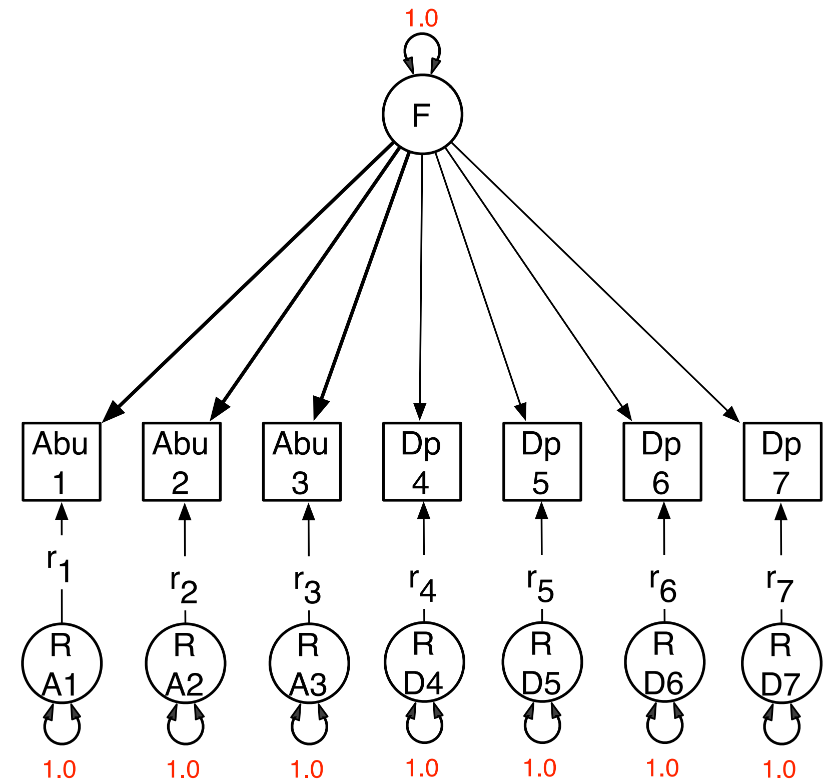
Jeannie-Marie S. Leoutsakos, PhD, MHS[1], **Peter P. Zandi, PhD, MHS**[2], **Karen Bandeen-Roche, PhD**[3], and **Constantine G. Lyketsos, MD, MHS**[1,2]

Factor Mixture Model

Class 1
probability
p



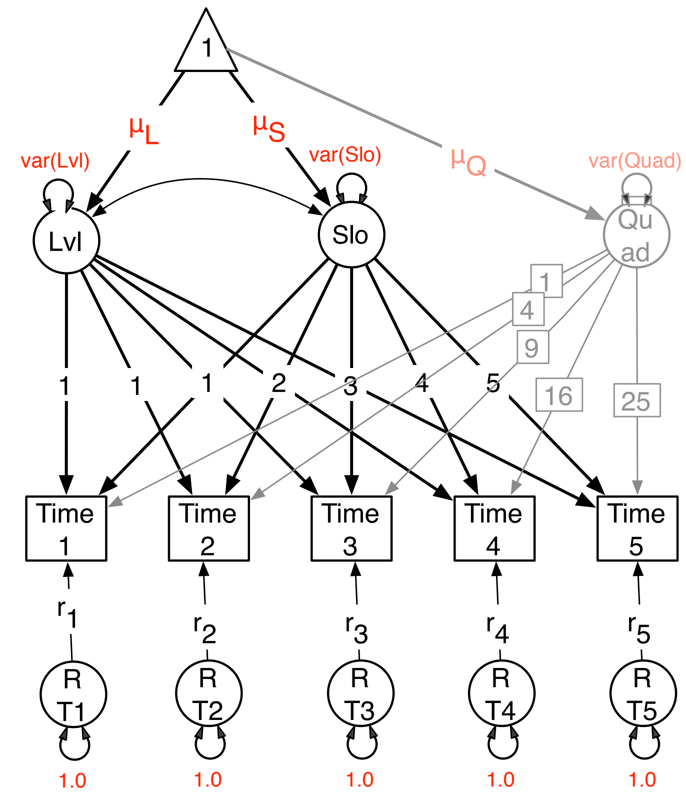
Class 2
probability
(1-p)



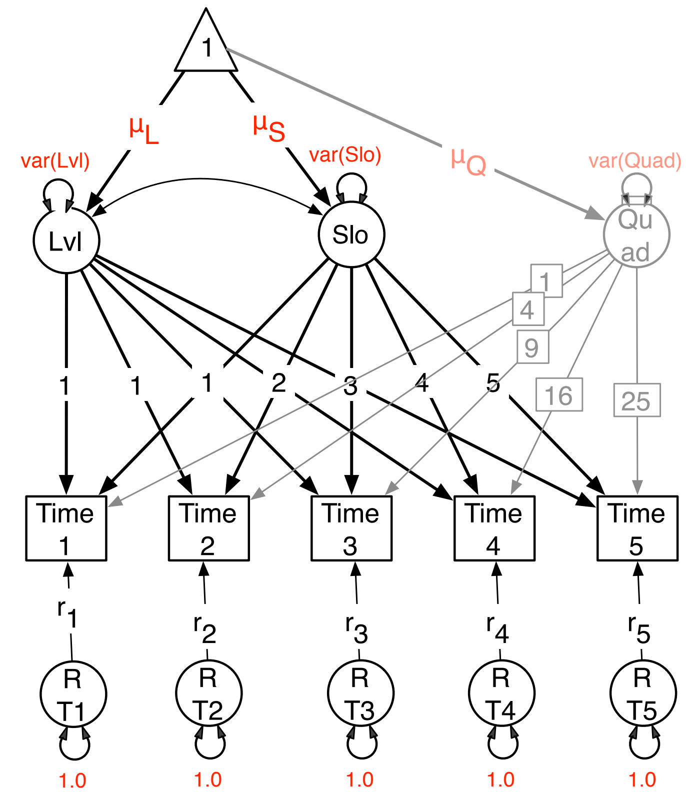
Very
Expensive!

Growth Curve Mixture

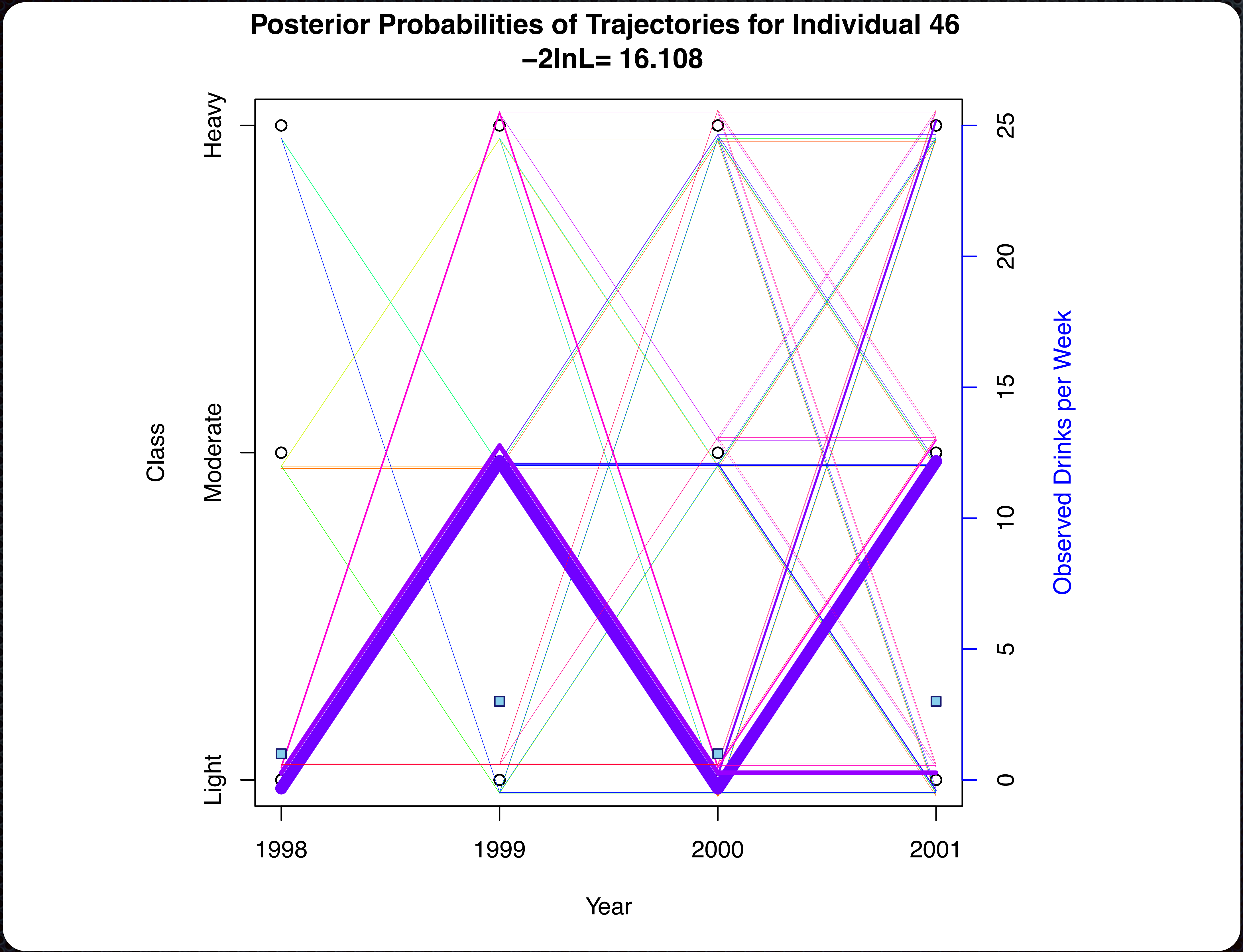
Class 1
probability
p



Class 2
probability
(1-*p*)

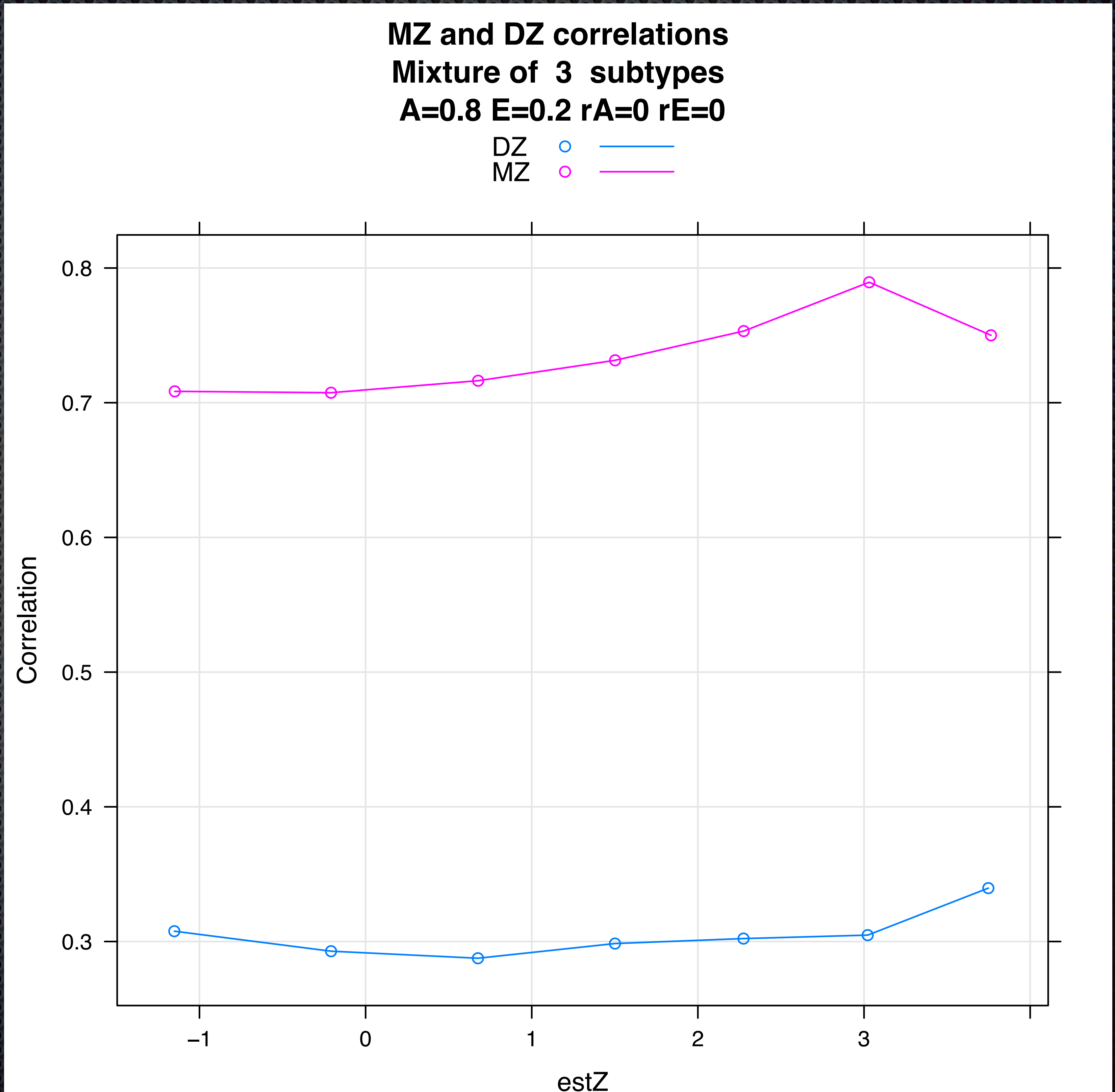


Regime Switching Model



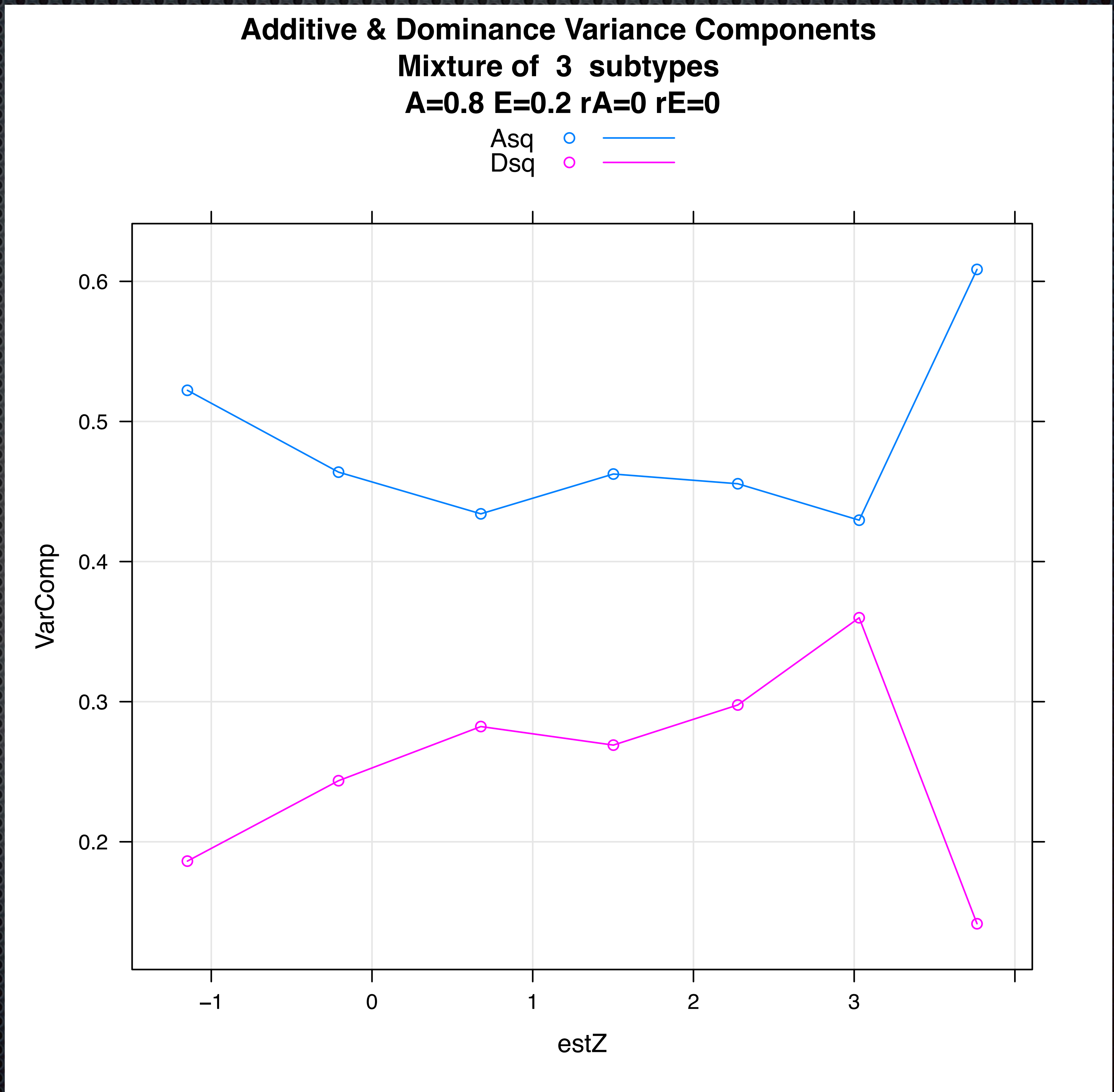
Occult Heterogeneity

- Suppose >1 uncorrelated heritable phenotypes
- Generate data, equal proportions
- Estimate correlations in mixture sample
- Vary threshold
- Induces VD



Occult Heterogeneity

- VA underestimated
- VD overestimated
- More effect the rarer the disorder
- Effect attenuates with genetic correlation between subtypes



Obligate Missingness

- Estimating correlation between Stem and Probe
  - 3+ categories of Stem and at least 2 lead to probe
  - 2 binary Stem items and endorsing either or both = probe
  - Binary Stem but collected from relatives who correlate < 1
- Do not mark missing probes as zero! Usually causes inflated item correlations

Obligate Missingness

- Stem: Have you ever used cocaine? 0/1/2
- Probe: Was it difficult to cut down or quit?
- Probe items are MAR conditional on Stem being 1 or 2
- WLS but not ML drastically attenuate correlation estimate
- Must code probes as missing!

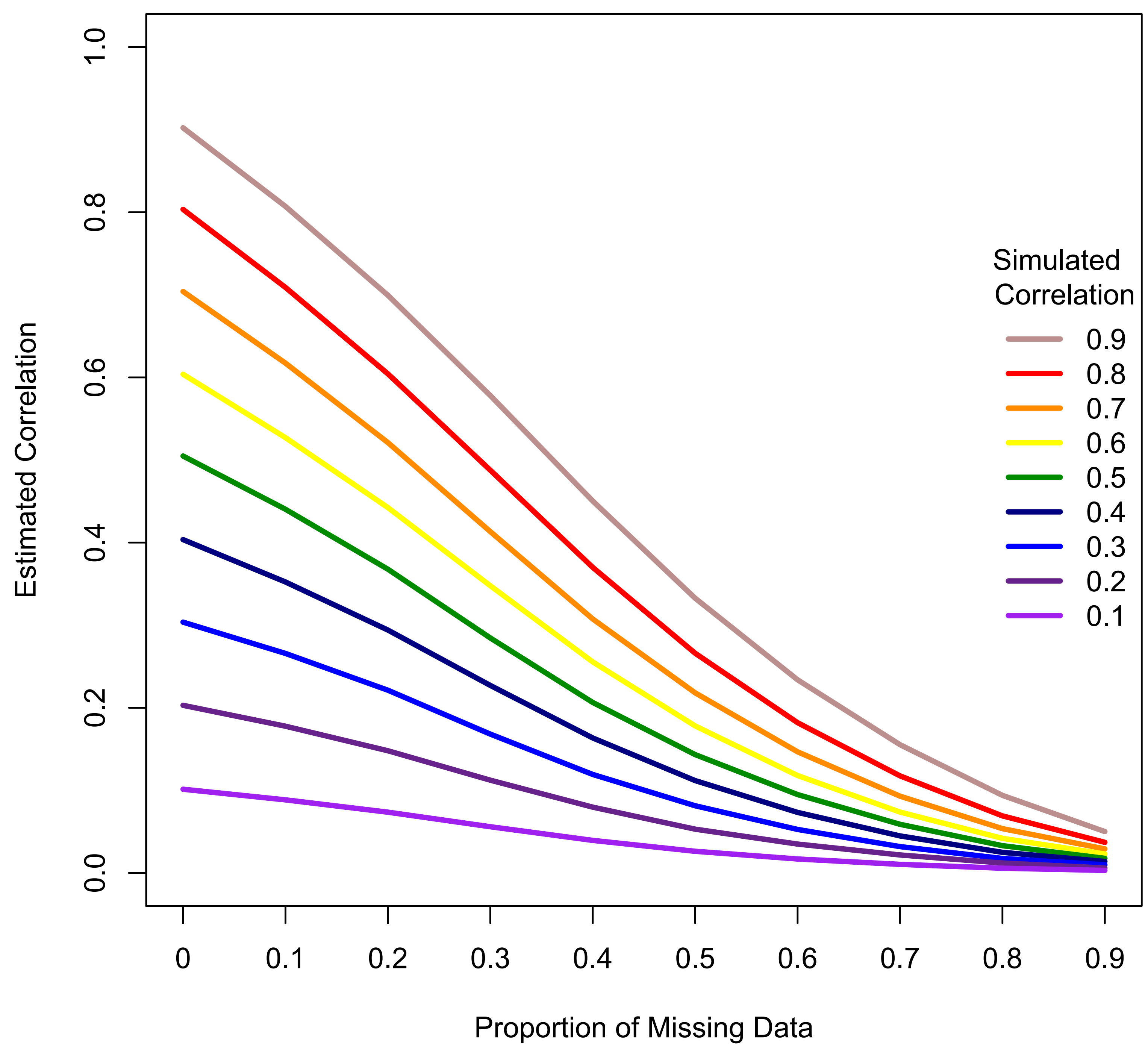
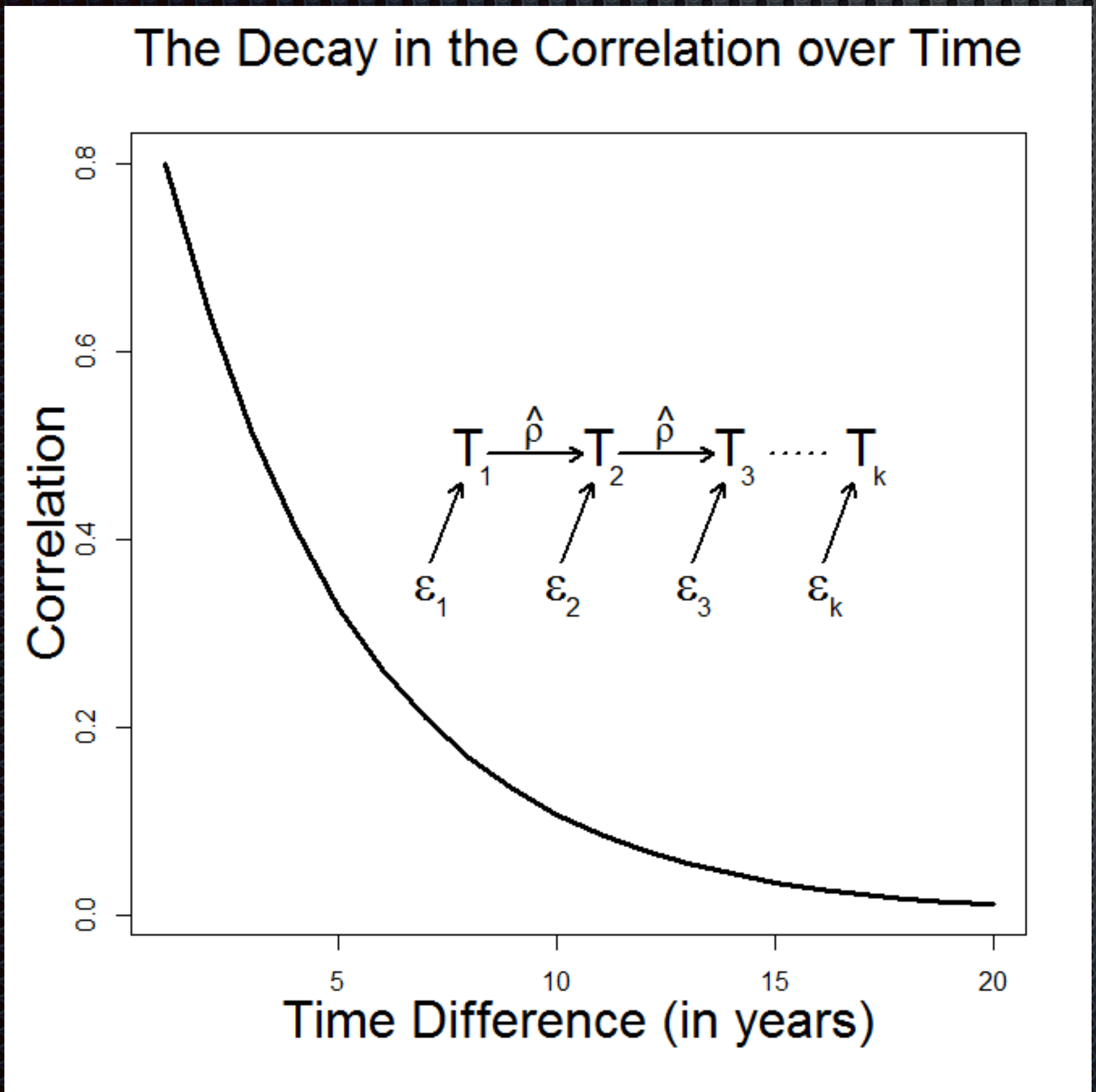


Figure 2: **Attenuation of the estimated correlation using WLS based on the level of MAR missingness.**

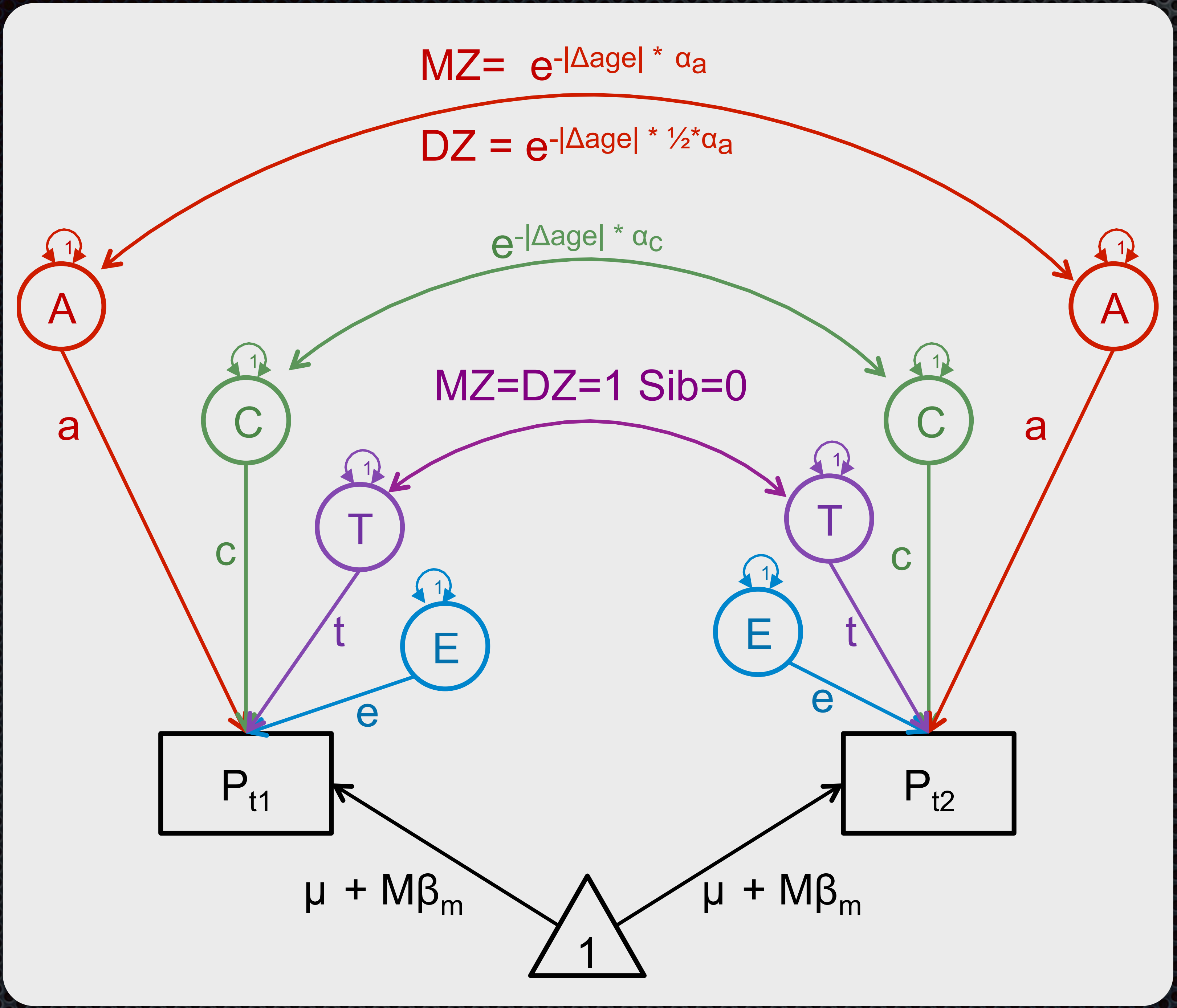
Genetic Heterogeneity with Age/Cohort

- Neuroticism within-person .6 correlation over 10 years
- Twin studies show rG < 1 over time
- Expressed genetic factors change during development
- Substance Use

Different age, different genes?



Age-Related Decay of Correlation



Verhulst, B., Eaves, L. J., and Neale, M. C. (Jul 2014). Moderating the covariance between family member's substance use behavior. Behav Genet, 44(4):337–46.

$$Cov = Acov * e^{-|\Delta age| * \alpha_a} + Ccov * e^{-|\Delta age| * \alpha_c} + Tcov$$

Application

Virginia 30,000 Data on Smoking

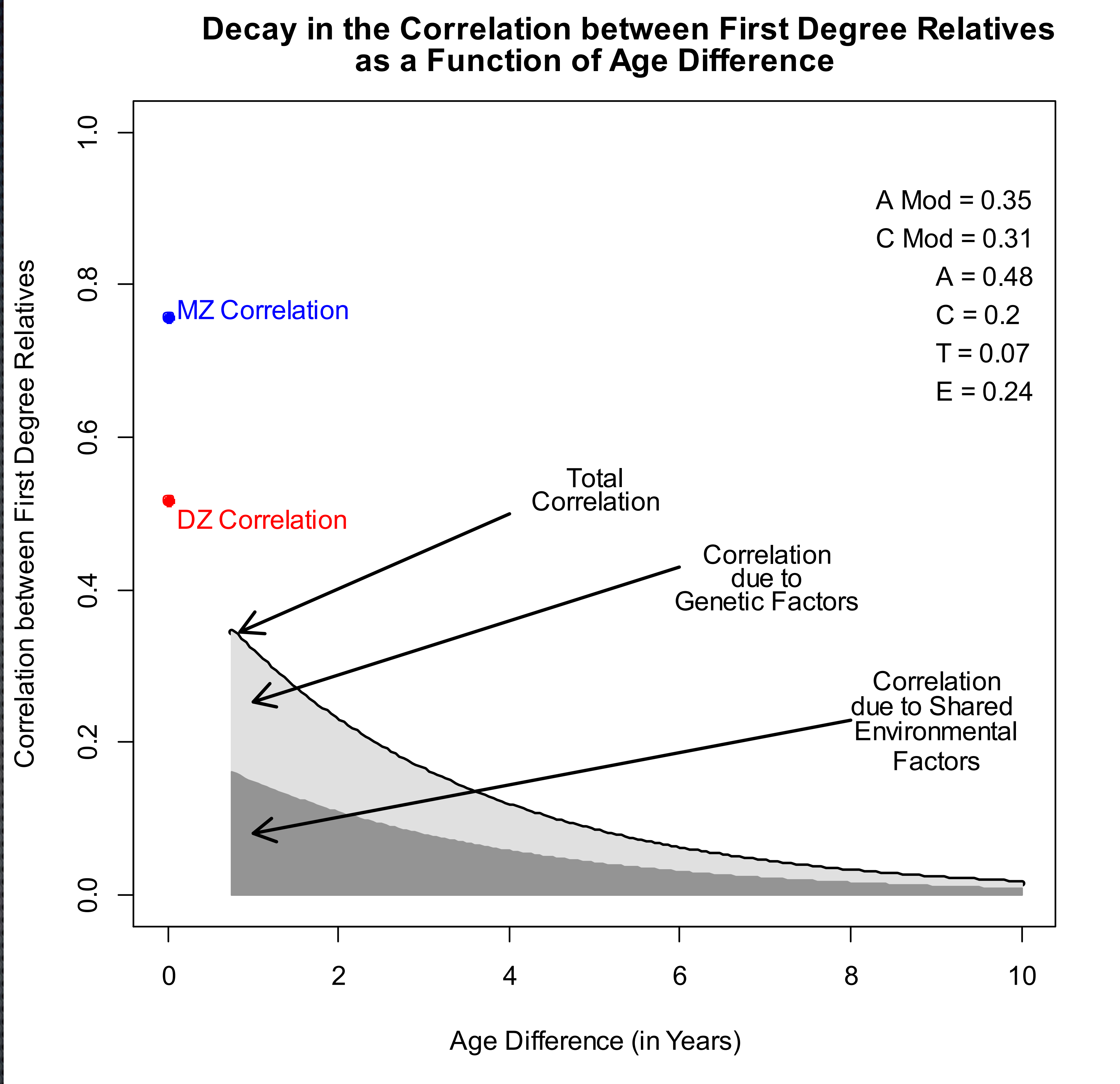
Twins, their parents, spouses, sibs and children

Twins only here, N=14,763

Crude smoking measure (1980s)

(1) never smoked, (2) used to smoke but gave it up, (3) smoked on and off, (4) smoked most of his/her life.

Strong evidence of decay with age difference



Future Directions

- Use “GCTA” Genetic relatedness matrices in place of close family relatives
  - Technical challenges, invert 20k x 20k matrices or larger
- Extend tests for direction of causation with combined twin and Mendelian Randomization model
- Dynamical models for high density repeated measures

Acknowledgements

NIDA R01 DA-18673

NIDA R25 DA-26119

NIMH R25 MH-19918

Study participants

Workshop participants

OpenMx Team

- VIPBG
- Brad Verhulst
- Shaunna Clarke
- Hermine Maes
- Steve Aggen
- Ken Kendler
- Computers