# Improved Measurement Modeling and Regression with Latent Variables

Karen Bandeen-Roche
Department of Biostatistics
Johns Hopkins University

**Department of Biostatistics and Computational Biology**
**University of Rochester**
**April 3, 2008**

# Acknowledgements

- Co-authors for Part I: Howard Chilcoat, Naomi Breslau
- Co-authors for Part II: Janne Petersen, Klaus Larsen
- Co-authors for Part III: Jeannie-Marie Leoutsakos, Jing Ning
- Thanks for data and collaboration: Sheila West, Linda Fried
- Funding: National Inst. on Aging, National Inst. of Mental Health, Johns Hopkins Older Americans Independence Center

# Introduction: Statistical Problem

- Observed variables (i=1,...,n): $Y_i$=M-variate; $x_i$=P-variate

- Focus: response (Y) distribution = $G_{Y|x}(y|x)$; x-dependence

- Modeling issue: flexible or theory-based?
  - — Flexible: $g_m(E[Y_{im}|x_i]) = f_m(x_i)$, m=1,...,M
  - — Theory-based:
    - > $Y_i$ generated from <u>latent</u> (underlying) $U_i$:

      $$F_{Y|U,x}(y|U=u,x;\pi) \qquad \textit{(Measurement)}$$

    - > Focus on distribution, regression re $U_i$:

      $$F_{U|x}(u|x;\beta) \qquad \textit{(Structural)}$$

    - > Overall, hierarchical, model:

      $$F_{Y|x}(y|x) = \int F_{Y|U,x}(y|U=u,x)\,dF_{U|x}(u|x)$$

# Motivation
# The Debate over Mixture and Latent Variable Models

- **In favor**: they
  - — acknowledge **measurement problems:** errors, differential reporting
  - — **summarize** multiple measures **parsimoniously**
  - — operationalize **theory**
  - — describe population **heterogeneity**

- **Against**: their
  - — **modeling assumptions** may determine scientific conclusions

  - — **interpretation** may be ambiguous
    - > nature of latent variables?
    - > comparable fit of very different models
    - > seeing is believing

## Possible Approaches to the Debate

- Argue advantages of favorite method

- Hybrid approaches:
  — Parallel analyses (e.g. *Bandeen-Roche et al. AJE 1999*)
  — Marginal mean + LV-based association
  (e.g. *Heagerty, Biometrics, 2001*)

- Sensitivity analyses

- **“Popperian”**

  **— Pose parsimonious model**

  **— Learn how it fails to describe the world**

# Outline

- Modeling and estimation framework

- Specifying the target of estimation

  — ***Supposing that the target uniquely exists ...***
  > Strategy for delineating it
  > Validity of the strategy

  — Application: Post-traumatic Stress Disorder

- Development and subsequent use of latent variable “indices”

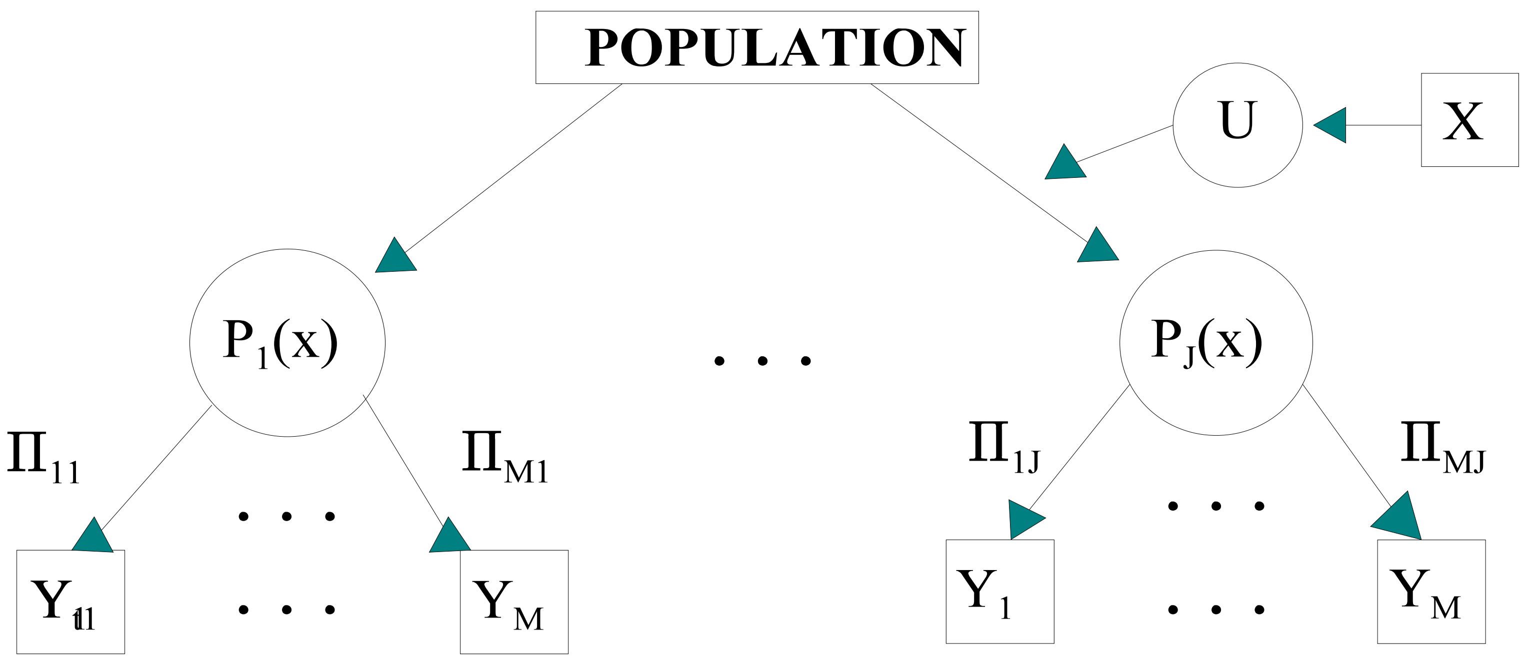
  — Application: Functioning and vision in older adults

- Refocusing: Methodology to counterbalance competing assumptions

## Application: Post-traumatic Stress Disorder Ascertainment

- PTSD
  - — Follows a qualifying traumatic event
    - > ***This study***: ***personal assault***, ***other personal injury/trauma***, ***trauma to loved one***, ***sudden death of loved one*** = "x", along with gender
  - — Criterion endorsement of symptoms related to the event ⇒ diagnosis
    - > Binary report on 17 symptoms = "Y"
- A recent study (Chilcoat & Breslau, *Arch Gen Psych*, 1998)
  - — Telephone interview in metropolitan Detroit
  - — n=1827 with a qualifying event
  - — Analytic issues
    - > Nosology
    - > Does diagnosis differ by trauma type or gender?
    - > *Are female assault victims particularly at risk?*

# Model 1
## Latent Class Regression




> $P_j(x) = \Pr\{U = j|x\}$
> $\pi_{mj} = \Pr\{Y_m=1|U = j\}$


*References: Dayton & Macready 1988, van der Heidjen et al., 1996; Bandeen-Roche et al., 1997*

# Latent Class Regression (LCR) Model

- **Model**:

$$f_{Y|x}(y|x) = \sum_{j=1}^{J} P_j(x,\beta) \prod_{m=1}^{M} \pi_{mj}^{y_m}(1-\pi_{mj})^{1-y_m}$$

- **Structural model assumption** : $[U_i|x_i] = Pr\{U_i=j|x_i\} = P_j(x_i,\beta)$
  — $RPR_j=Pr\{U_i = j|x_i\}/Pr\{U_i = J|x_i\}$; $j=1,\ldots,J$

- **Measurement assumptions** : $[Y_i|U_i]$
  — conditional independence
  — nondifferential measurement
  > ***reporting heterogeneity unrelated to measured, unmeasured characteristics***

- **Fitting**: ML w EM; robust variance (e.g. *Muthén & Muthén 1998, M-Plus)*

- *Posterior* latent outcome info: $Pr\{U_i=j|Y_i,x_i;\theta=(\pi,\beta)\}$

# Methodology

## Delineating the Target of Measurement

- **Fit an initial model**: ML, Bayes, etc.

- **Obtain *posterior* latent outcome** info — e.g. $f_{U|Y,x}(u|Y,x;\theta)$
  — This talk: empirical Bayes

- RANDOMLY generate "empirical LVs," $V_i$, according to $f_{U|Y,x}(u|Y,x;\hat{\theta})$

- Analyze $V_i$ AS $U_i$ (accounting for variability in first-stage estimation)

- Estimate measurement structure through empirical analysis of $Y_i|V_i,x_i$

# Methodology

## Properties “whatever” the True Distribution

- Under Huber (1967)-like conditions:

  — <u>Asymptotically</u>:

  > Randomization imposes limiting hierarchical model, except [Y|V,x] arbitrary (and specifiable)

  i.e. ***underlying variable distribution has an estimable interpretation even if assumptions are violated***

  > No bias in substituting $V_i$ for $U_i$.

  i.e. ***regression of $V_i$ on $x_i$ and model-based LV regression eventually equivalent***

# Methodology
## More formal statement

- Under Huber (1967)-like conditions:

—$(\hat{\beta},\hat{\pi})$ converge in probability to limits $(\beta^*,\pi^*)$.

—$Y_i$ asymptotically equivalent in distribution to $Y^*$, generated as:

i) Generate $U_i^*$ — distribution determined by $(\beta^*,\pi^*)$, $G_{Y|x}(y|x)$;

ii) Generate $Y^*$—distribution determined by $(\beta^*,\pi^*)$, $G_{Y|x}(y|x)$, $U_i^*$

— $\{\Pr[Y_i \leq y|V_i,x_i], i=1,2,...\}$ converges in distribution to $\{\Pr[Y_i^* \leq y|U_i^*,x_i], i=1,2,...\}$, for each supported y.

— $V_i$ converges in distribution to $U_i^*$.

# PTSD Study: Descriptive Statistics

| Gender | Trauma Type: percentage distribution | | | | n |
|---|---|---|---|---|---|
| | *Personal Assault* | *Other Injury* | *Trauma to loved one* | *Sudden death* | |
| *Male* | 14.2 | **37.7** | 26.9 | 21.3 | 964 |
| *Female* | 14.3 | 26.3 | **32.2** | **27.2** | 863 |
| **Total** | 14.2 | 32.3 | 29.4 | 24.1 | 1827 |

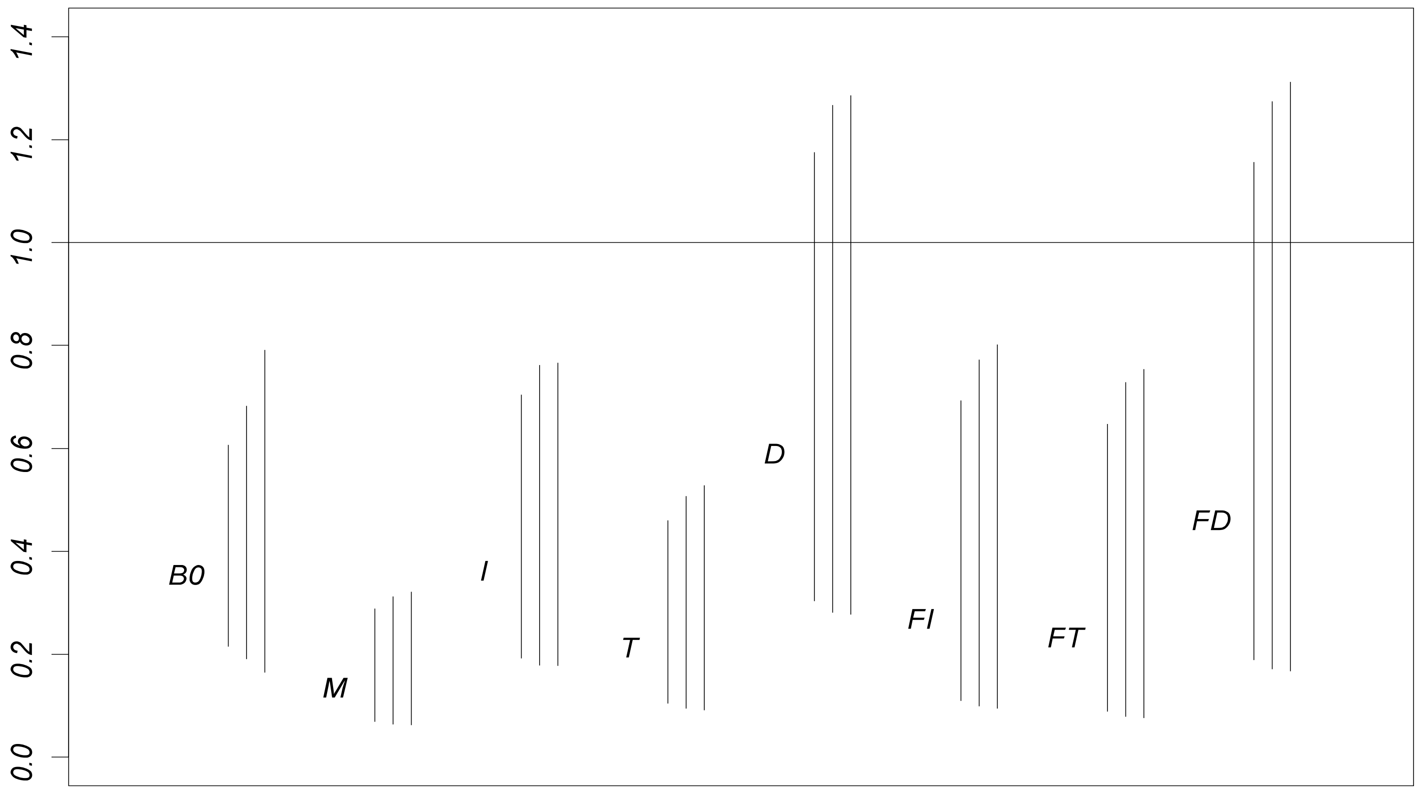
- PTSD symptom criteria met: 11.8% (n=215)
  - — By gender: 8.3% of men, 15.6% of women
  - — By trauma: *assault (26.9%), sudden death (14.8%), other injury (8.1%), trauma to loved one (6.0%)*
  - — Interactions: female x assault (↑), female x other (↓)
  - — Criterion issue? 60% reported symptoms short of diagnosis

# Latent Class Model for PTSD: 9 items

| SYMPTOM CLASS | SYMPTOM (prevalence) | SYMPTOM PROBABILITY (π) | | |
|---|---|---|---|---|
| | | Class 1 - NO PTSD | Class 2 - SOME SYMPTOMS | Class 3 - PTSD |
| **RE-EXPERIENCE** | Recurrent thoughts (.49) | .20 | **.74** | .96 |
| | Distress to event cues (.42) | .12 | **.68** | .88 |
| | Reactivity to cues (.31) | .05 | **.51** | .77 |
| **AVOIDANCE/ NUMBING** | Avoid related thoughts (.28) | .08 | .37 | .75 |
| | Avoid activities (.24) | .05 | .34 | .66 |
| | Detachment (.15) | **.01** | **.14** | **.64** |
| **INCREASED AROUSAL** | Difficulty sleeping (.19) | .02 | .18 | .78 |
| | Irritability (.21) | .02 | .22 | .83 |
| | Difficulty concentrating (.25) | .03 | .30 | .89 |
| **MEAN PREVALENCE-BASELINE** | | .52 | .33 | .14 |

[Omitted: nightmares, flashback; **amnesia, ↓interest, ↓affect, short future;** hypervigilance, startle]

*Odds and Relative Odds, with 95% Confidence Intervals*

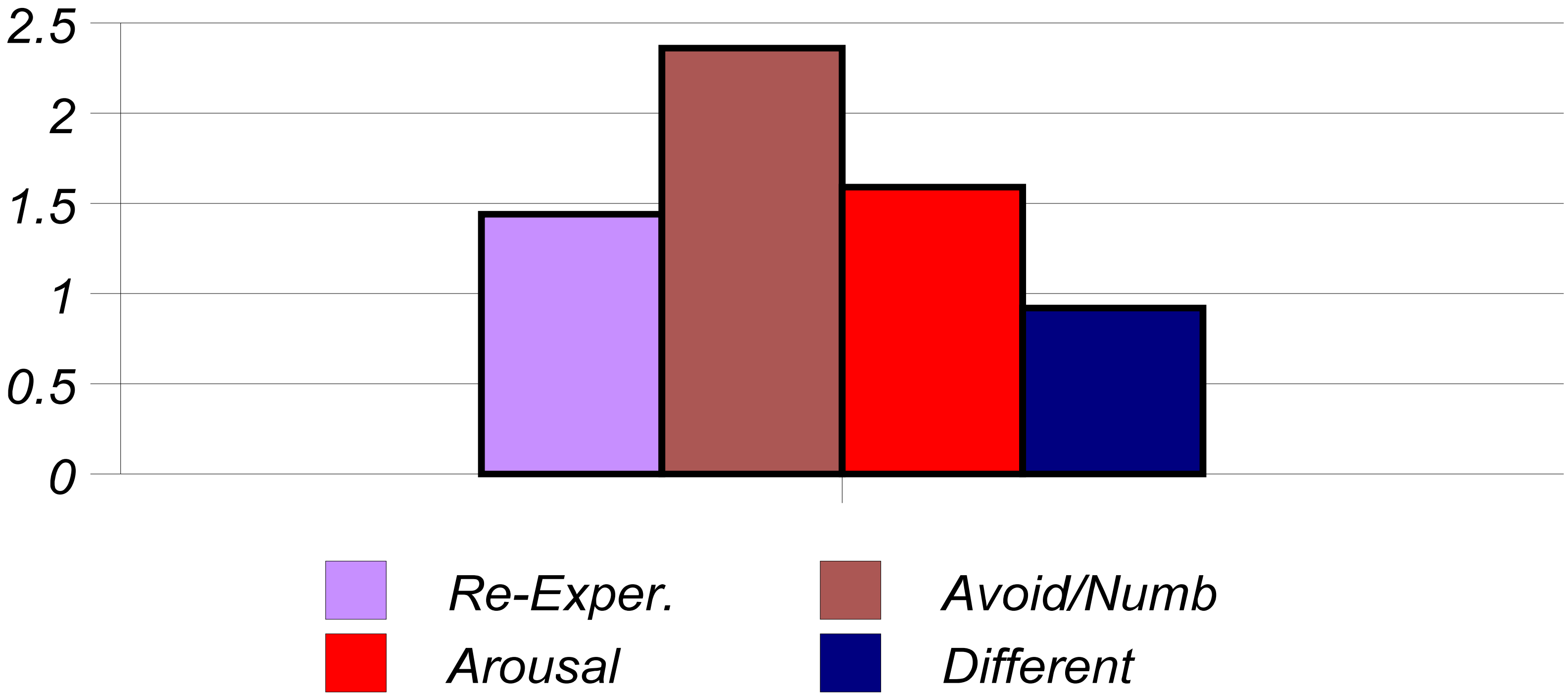


# PTSD: DIAGNOSIS, LCR MEASUREMENT MODEL

- Method: Regress item responses on covariates "controlling" for class
  — For simplicity: non-assaultive traumas merged into "other trauma"

| Variable | Odds Ratio or Interaction Ratio (CI) | By-item Odds Ratio MODEL 2 |
|---|---|---|
| Female | 1.07 (0.93,1.22) | 1.07 (0.93,1.22) |
| Trauma =other than assault (recur.) | **3.19 (1.89,5.40)** | **3.19 (1.89,5.40)** |
| Cue distress x other trauma | **0.18 (0.09,0.38)** | **0.58 (0.36,0.92)** |
| Cue reactivity x other trauma | **0.14 (0.07,0.28)** | **0.44 (0.27,0.72)** |
| Avoid thoughts x other trauma | **0.21 (0.11,0.41)** | ***0.68 (0.44,1.05)*** |
| Avoid activities x other trauma | **0.11 (0.05,0.22)** | **0.35 (0.21,0.58)** |
| Detachment x other trauma | **0.27 (0.13,0.58)** | 0.88 (0.51,1.49) |
| Difficulty sleep x other trauma | **0.43 (0.21,0.90)** | 1.37 (0.78,2.42) |
| Irritability x other trauma | **0.28 (0.13,0.61)** | 0.91 (0.52,1.59) |
| Concentration x other trauma | 0.73 (0.36,1.47) | **2.33 (1.35,4.03)** |

## Diagnosis:  Conditional Independence (Pairwise ORs)




| | |
|---|---|
| Re-Exper. | 1.44 |
| Avoid/Numb | 2.36 |
| Arousal | 1.59 |
| Different | 0.92 |

# Summary
## PTSD Analysis

- The analysis hypothesizes that PTSD is
  - — a syndrome comprising **unaffected**, **subclinically affected**, and **diseased** subpopulations of those suffering traumas
  - — reported homogeneously within subpopulations
- The hypotheses are consistent with current diagnostic criteria
- Gender x type interactions: are strongly indicated
  - — Female assault victims at particular risk
  - — ... given the subpopulations defined by the model

# Summary
## PTSD Analysis

- Symptoms appeared differentially sensitive to different traumas

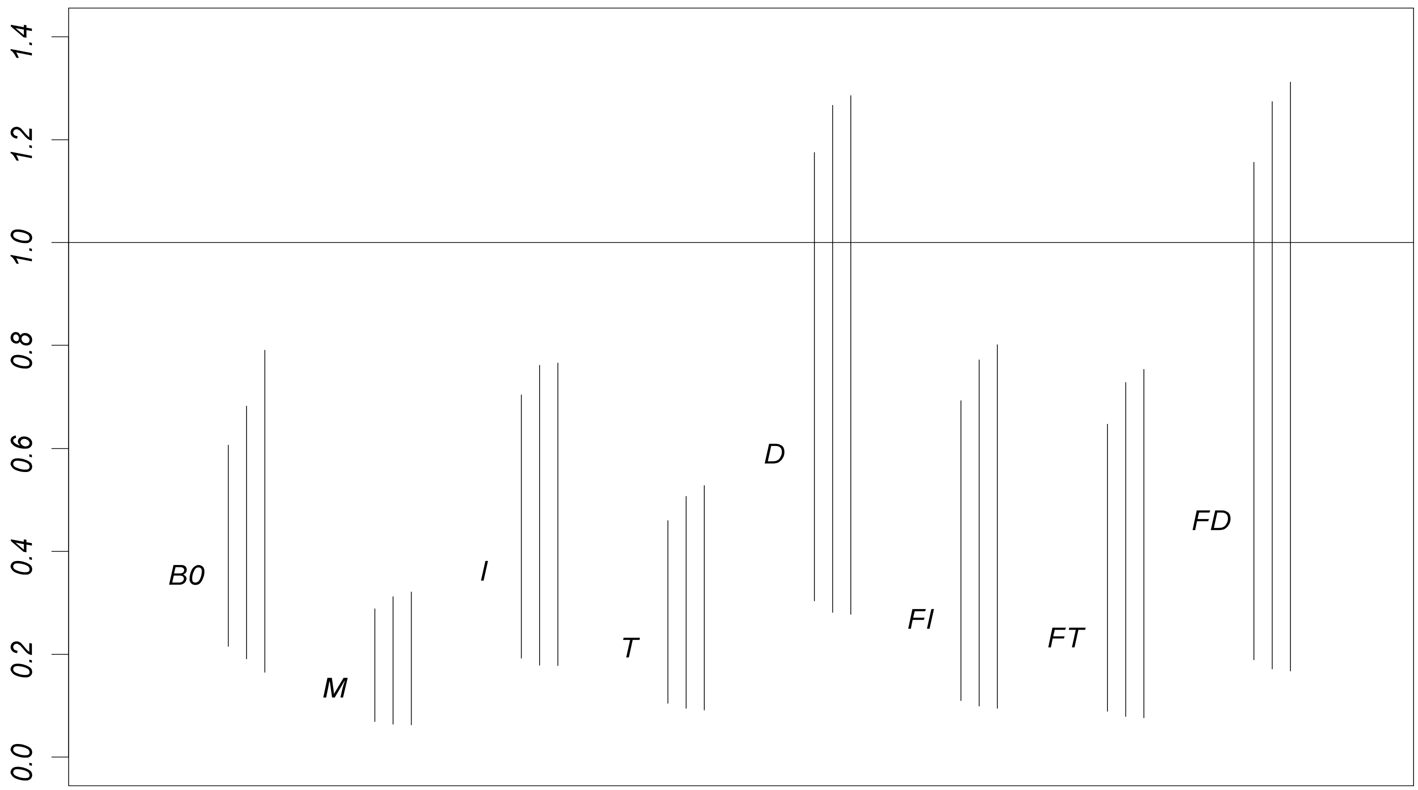
  Within classes: those who had a non-assaultive trauma were

  — **less prone** to report distress to cues, reactivity to cues, avoiding thoughts, & avoiding activities

  — **more prone** to report recurrent thoughts & difficulty concentrating

- Concern: Current criteria may better detect psychiatric sequelae to assault than to traumas other than assault

Odds and Relative Odds, with 95% Confidence Intervals



# Latent Variable Scaling
# A Three-Stage Approach

- **<u>Step 1</u>**: Fit full latent variable measurement model $\Rightarrow \hat{\pi}$
  - — For now: Non-differential measurement
- **<u>Step 2</u>:** Obtain predictions $O_i$ given $\hat{\pi}$, $Y_i$
- **<u>Step 3</u>**: Obtain $\hat{\beta}$ via regression of $O_i$ on $x_i$
- **<u>Step 4 (rare)</u>**: Fix inferences to account for uncertainty in $\hat{\pi}$

# Latent Variable Scaling (obtaining $O_i$)
## What do we know?

---

- **<u>Predominant work</u>:** Latent Factor models

  — $U \sim \text{Normal}$; $[Y|U] \sim \pi U + \epsilon$, $\epsilon \sim N(0,\Sigma)$

  — **Three scaling methods**

  > **Ad hoc**

  > **Posterior mean**: $O_i$ as $E[U_i|O_i, \hat{\pi}]$

  > **"Bartlett" method**: Weighted least squares, $U_i$ "fixed"

  $Y_i = \hat{\pi} U_i + \epsilon_i$, $\epsilon_i \sim N(0, \hat{\Sigma})$; $O_i$ as WLS model fit for $U_i$

  — **In Step 3, Bartlett scores yield consistent** $\hat{\beta}$; others don't

## Latent Variable Scaling (obtaining $O_i$)
### What do we know?

- **Latent Class models**
  - — **Two scaling methods**

    > **Posterior class assignment**
    - Modal or as "pseudo-class": single or multiple

    > **Posterior probability estimates**:

$h_i = f_{U|Y}(u|Y; \hat{\pi})$; $O_i = h_i$ (logit link) or logit($h_i$) or weighted

— **In Step 3, all are biased for** $\hat{\beta}$

— A **correction**: Croon, *Lat Var & Lat Struct Mod*, 2002
Bolck et al., *Political Analysis*, 2004

## Latent Variable Scaling (obtaining $O_i$)
### A new proposal

- **Motivation**: Bartlett method
  - — $[Y|U]$ ~ product Bernoulli, $p = \pi S(U)$

    > Y, p: Mx1 vectors (**outcomes**)

    > $\pi$: MxJ matrix of conditional probabilities (design matrix)

    > S(U): Jx1 vector with jth element = $\mathbf{1}\{U=j\}$ ("coeffs")

  - — Proposed **Step 2**: GLM of $Y_i$ on $\hat{\pi}$ with **linear** link, Bernoulli family; $O_i = \hat{S}_i$

  - — ML for GLM can be written as IRWLS
  - — **A shortcut**: $O_i = \hat{S}_i$ via **ordinary** least squares; COP score

# Simulation Study

- Basic template: 2 classes; $\pi = \begin{pmatrix} \tau & 1-\tau \\ & \vdots & \\ \tau & 1-\tau \end{pmatrix}$

  —**2 measurement scenarios**: "**Precise**"–$\tau$=0.10; "**Imprecise**"–$\tau$=0.30

- M=4, 8

- n=500, 1000

- 2 covariates; $\beta_0 = 0$ ; $\beta_1 = \beta_2 = 0.5$

- Lots of secondary simulations to compare COP scores, full LV

# COP Scoring
# Theory

- Proposed **Step 3**: GLM of O on x with **gen. logit** link, Normal family

- **Punch line:** **In Step 3, COP scores yield consistent $\hat{\beta}$.**

- **Basic ideas**

— **If $\pi$ were known**: OLS yields unbiased estimator of $\begin{pmatrix} Pr\{U_i = 1\} \\ \vdots \\ Pr\{U_i = J\} \end{pmatrix}$

$$> \begin{pmatrix} Pr\{U_i = 1\} \\ \vdots \\ Pr\{U_i = J\} \end{pmatrix} = \begin{pmatrix} P_1(x_i, \beta) \\ \vdots \\ P_J(x_i, \beta) \end{pmatrix}, \text{ all i}, \Rightarrow \hat{\beta}_{COP} \xrightarrow{p} \beta$$

— $\hat{\pi} \xrightarrow{p} \pi$ (marginalization, ML); then, uniform integrability

# Simulation Study Results

| Method | Precise, m=4, n=500 | | | Imprecise, m=4, n=1000 | | | Imprecise, m=8, n=1000 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $E\hat{\beta}_1$ | $SE_{rat}$ | Cov | $E\hat{\beta}_1$ | $SE_{rat}$ | Cov | $E\hat{\beta}_1$ | $SE_{rat}$ | Cov |
| Modal class | 0.48 | 1.00 | 0.95 | 0.30 | 0.96 | 0.68 | 0.37 | 1.03 | 0.83 |
| Pseudo-class | 0.47 | 0.98 | 0.95 | 0.24 | 0.97 | 0.50 | 0.33 | 1.03 | 0.76 |
| Posterior-GLM | 1.66 | 0.98 | 0.59 | 0.33 | 0.96 | 0.71 | 0.62 | 0.98 | 0.92 |
| Croon corrected | 0.51 | NA | NA | 0.49 | NA | NA | 0.47 | NA | NA |
| COP score | 0.51 | 0.97 | 0.95 | 0.51 | 0.98 | 0.96 | 0.49 | 1.00 | 0.94 |
| LCR | 0.51 | 0.99 | 0.95 | 0.52 | 0.98 | 0.96 | 0.49 | 1.02 | 0.95 |

- n=500 vs 1000, m=8: negligible difference
- Power = slightly highest for LCR; others = ~ comparable except pseudo
  - Relative efficiency re LCR: $\geq 0.89$

## Simulation Study
## COP Score Performance in Secondary Runs

---

- Findings similar in many cases:
  — 3 classes
  — $\beta_0 \neq 0$, different $\beta_1$
  — different measurement models
  — continuous versus binary x

- Multiple (4) covariates

  — Accuracy of mean model estimation maintained

  — Accuracy of standard errors compromised

  > For moderate $|\beta_1|$: coverages ~ within 0.02 of 0.95

  > With large $|\beta_1|$: coverages as low as 0.83

# Application

## IADL Functioning in the Salisbury Eye Evaluation (SEE) Study

- **Study**: Salisbury Eye Evaluation (SEE; West et al. 1997)
  — Representative of community-dwelling elders
  — n=2520; 1/4 African American
  — This talk: A convenience sample of n=1329

- **Question of interest**: Is worse vision associated with worse IADL functioning independently of age (and sex)?

  — IADL (Y): Indicators of difficulty **shopping**, **preparing meals**, **doing light housework**, and **using the phone**

  — Vision (primary X): Visual acuity (logMAR)

# Application
## Findings

- Two class model (questionable fit)

| Coefficient | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | LCR | COP | LCR | COP |
| Intercept | -3.17 (-3.61,-2.73) | -3.12 (-3.51,-2.73) | -2.91 (-3.44,-2.34) | -3.02 (-3.47,-2.57) |
| Vision | 2.05 ( 1.33, 2.76) | 2.15 ( 1.72, 2.59) | 2.00 ( 1.21, 2.78) | 2.11 ( 1.68, 2.55) |
| Age (yr) | 0.75 ( 0.21, 1.29) | 0.72 ( 0.28, 1.17) | 0.72 ( 0.17, 1.26) | 0.71 ( 0.27, 0.15) |
| Sex | NA | NA | -0.68 (-1.34,-0.03) | -0.17 (-0.63, 0.28) |

— Re green estimates: many other methods closer to LCR

## Summary

- What I delineated
  - — A philosophy
    - > Fit an ideal model
    - > Determine the nature of measurement achieved in fact
  - — Theory: On the nature of measurement
  - — Methodology: To implement the philosophy
  - — New work: On regression with latent variable indices; on compromise between potentially competing validation criteria
- Strengths / benefits
  - — Improved use / usefulness of latent variable models
  - — Improved accuracy of regression using latent class scores
  - — Allows some distrust of the data

# Discussion

- <u>A primary issue</u>: Why a hierarchical model at all?

  — PTSD: Why not DSM *Y*, delineate measurement properties?

  1) **Nosology**
     a. Central role of cond. independence, non-diff. measurement.
     b. Guidance in creating, say, three rather than two groups.

  2) **The quest for the “ideal”**
     a. Could have turned out that LCR much less subject to NDM, than DSM: i.e. issue with diagnostic criteria rather than items.

     b. <u>In fact</u>: LCR and DSM about equally subject to NDM

     c. Ultimate recommendation: DSM

## Discussion

- Beyond delineation of assumptions....

- Further work: Uniqueness of target
  — Delineation of plausible models
  — Displays, complicated models
  — ***Implication:*** Guidance on parsimony versus complexity

- Further work: Latent class scoring
  — Consistent inference
  — Case of differential measurement

- Further work: Big picture for validation compromise
  — How does measurement conform?
  — How should one determine the magnitude of the compromise?

- Why not be Bayesian?

# Implications

- More valid usage of latent variable modeling
- Provision of more clearly interpretable scales
- Improved delineation of health statuses and inference regarding etiology