On the Targets of Latent Variable Model Estimation

Karen Bandeen-Roche
Department of Biostatistics
Johns Hopkins University

Department of Mathematics and Statistics
Miami University
December 8, 2005

With thanks to: Howard Chilcoat, Naomi Breslau, Linda Fried

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

  — ***Unique existence of the target***

- Applications
  — Post-traumatic Stress Disorder
  — Basic task disability in older women

Application: Post-traumatic Stress Disorder Ascertainment

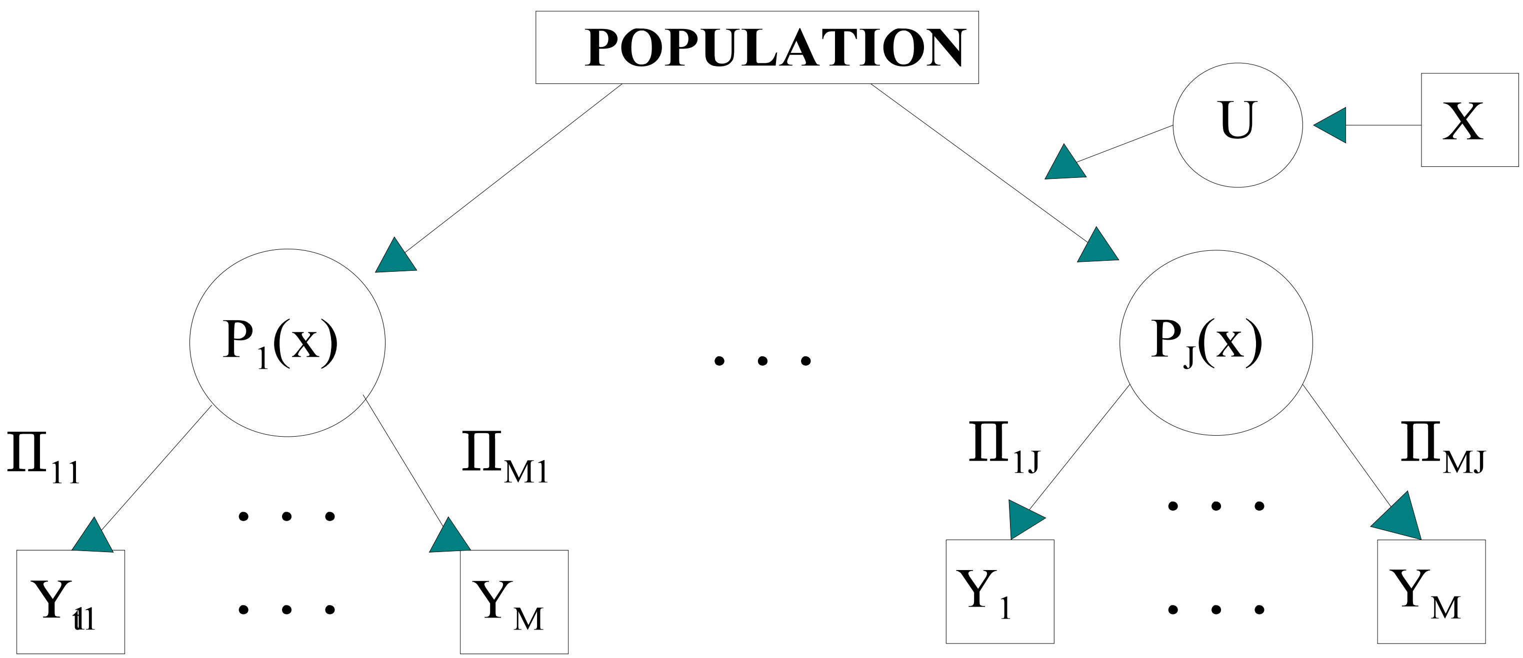
- PTSD
  — Follows a qualifying traumatic event
    > ***This study***: ***personal assault***, ***other personal injury/trauma***, ***trauma to loved one***, ***sudden death of loved one*** = "x", along with gender

  — 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



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

<u>References</u>: *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 (β^*,π^*).

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

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

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

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

  — V_i converges in distribution to U_i^*.

PTSD Study: Descriptive Statistics

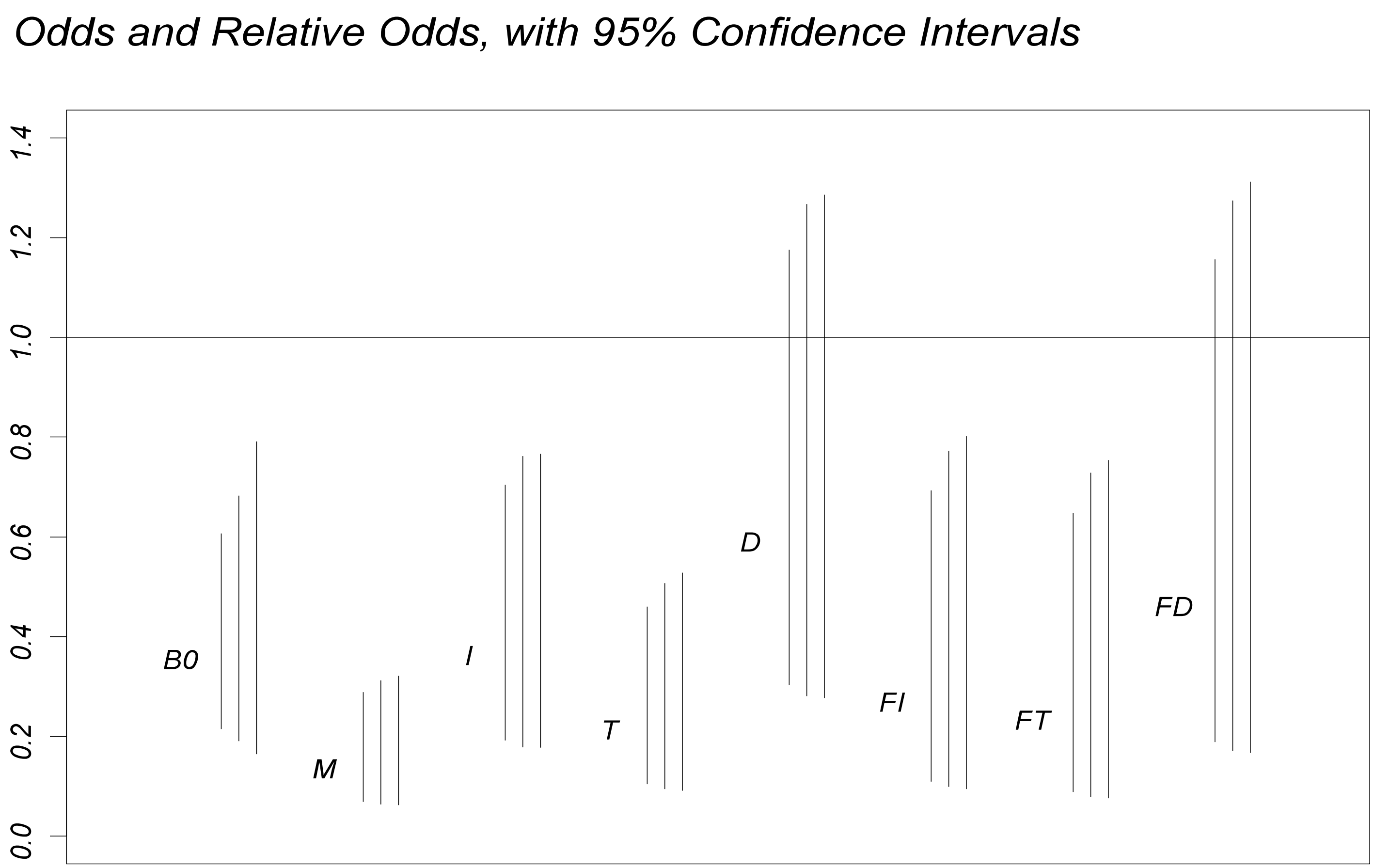
Gender	**Trauma Type: percentage distribution**				**n**
	Personal Assault	*Other Injury*	*Trauma to loved one*	*Sudden death*	
Male	14.2	**37.7**	26.9	21.3	964
Female	14.3	26.3	**32.2**	**27.2**	863
Total	14.2	32.3	29.4	24.1	1827

- 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]



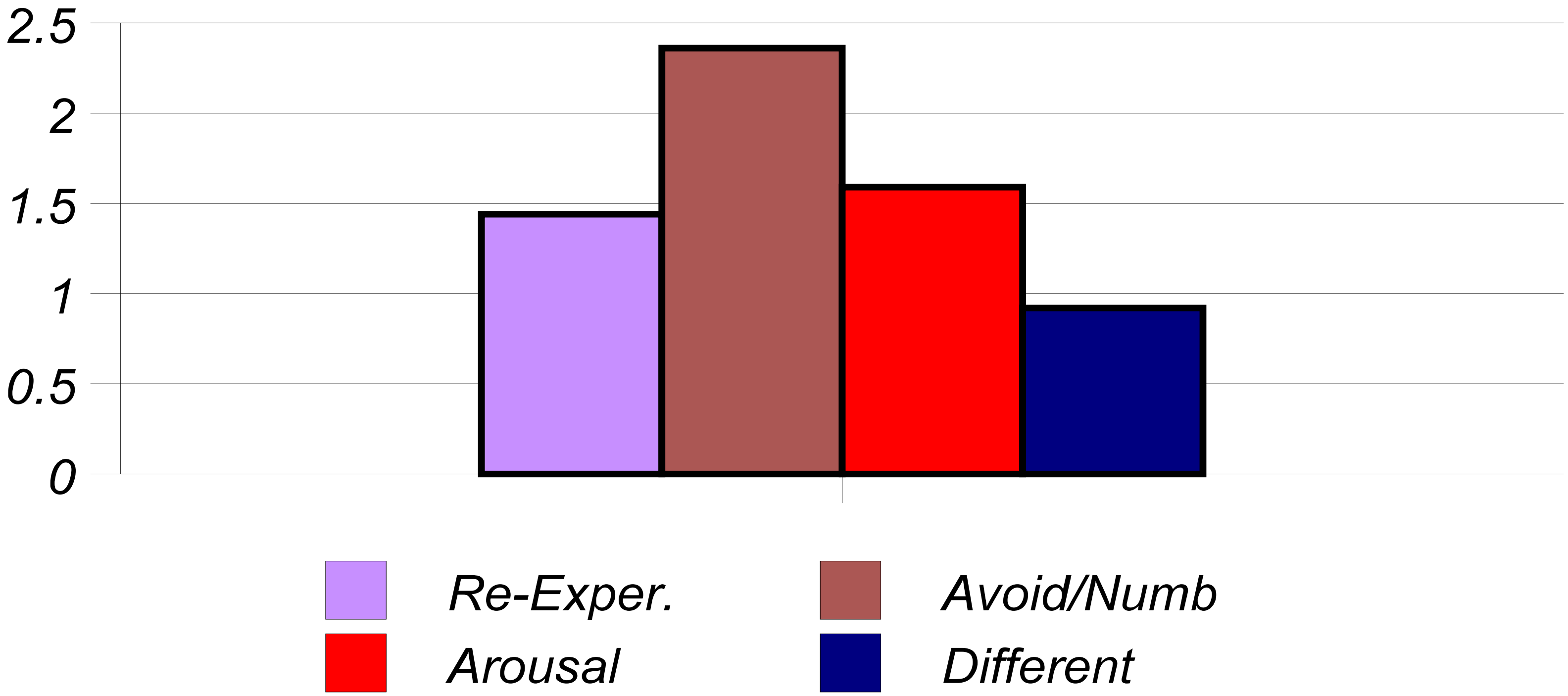
Odds and Relative Odds, with 95% Confidence Intervals
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
B0
M
I
T
D
FI
FT
FD

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

Characterization of the Target Parameters (β^*,π^*)

Huber (1967), Proc. 5th Berkeley Symposium

- Notation
  - — True distribution: $\{Y_1,\ldots,Y_n\}$ i.i.d. with $Y_i \sim \boldsymbol{f}^*{}_Y(y)$
  - — Model: $Y_i \sim \boldsymbol{f}_Y(y;\beta,\pi)$ ~ an ***LCA*** mass function
  - — Derivative operator: D_ϕ = gradient wrt (β,π)
- If (β^*,π^*) exist: they minimize Kullback-Leibler distance between $\boldsymbol{f}$ & $\boldsymbol{f}^*$
- When do (β^*,π^*) exist?
  - — Regularity conditions
  - — **Key**: $E_{f^*}[D_\phi \ln \boldsymbol{f}_Y(y;\beta,\pi)]$ has a unique 0

Existence of the Target Parameters (β^*,π^*)

Verification

- Two strategies
  - Theory
    - > Geometry (e.g. *Lindsay, Ann Stat., 1983*)
    - > Global identifiability
  - **Direct examination**
    - > $\sum_{l=1}^{2^M} [\boldsymbol{D_\phi \ln f_Y(y_l;u,v)} \,|_{\beta,\pi}]\; \boldsymbol{Pr\{Y=y_l\}}$ as a function of (β,π) (grid)
    - > $\boldsymbol{Pr\{Y=y\}} = \boldsymbol{f^*_{\mathrm{Y}}(\mathrm{y})}$ unknown; estimate by empirical $\hat{P}\boldsymbol{\{Y=y\}}$
- <u>A key aid</u>: substantive conceptual framework
  - Reduction of parameter space

Example: Self-reported Disability among Older Adults

- Import: Medicare funding weighs prevalence of self-reported disability
  — Recent report: disability decreasing *(Manton et al., 1998)*

- This talk: Basic functioning in The Women's Health and Aging Study
  — "Basic function" via "Do you have difficulty ..."
  > *bathing*, preparing *meals*, *dressing*, using the *toilet* (M=4)

  — 7 rounds every 6 months: n=1002 at baseline

  — Aims: disability prevalence trend + role of covariates, x_{it}

  — Potential failure to measure as intended: trust effect

- Why not a "harder" outcome than self-report?
  — Distinct dimension of health *(e.g., Jette, 1980)*
  — Increasingly a focus of interventions

Example
Women's Health and Aging Study

- Conceptual framework: Task hierarchy (*Fried et al., J Clin Epi, 1999*)
  — Difficulty ordering according to physiological demand

  — Basic functioning: bathing = most difficult; others = "parallel"

- **Idealized** conditional probabilities (πs):

Task	**Class 1**	**Class 2**	**Class 3**
Bathing	0	1	1
Preparing Meals	0	0	1
Dressing	0	0	1
Use toilet	0	0	1

Example
Women's Health and Aging Study

- <u>Conceptual framework</u>: Task hierarchy (*Fried et al., J Clin Epi, 1999*)
  — Difficulty ordering according to physiological demand

  — Basic functioning: bathing = most difficult; others = "parallel"

- Modeled conditional probabilities (πs):

Task	**Class 1**	**Class 2**	**Class 3**
Bathing	0	1	π_3
Preparing Meals	π_1	π_2	π_3
Dressing	π_1	π_2	π_3
Use toilet	π_1	π_2	π_3

> Constrained parameter space: $\pi_1 < .5$, $\pi_2 < .5$, $\pi_3 > .5$

Example: Uniqueness of Target Parameters

- First, a test case: $P_1 = P_2 = 1/3$; $(\pi_1,\pi_2,\pi_3) = (.1,.1,.9)$
- Measure of closeness to 0: Euclidean norm
  > 5-number summary: 0.00, 0.20, 0.33, 0.52, 0.75

- **Constrained** Grid (P,π) with 10 expected gradients closest to 0:

P_1	P_2	π_1	π_2	π_3	Norm
0.3333	0.3333	0.1000	0.1000	0.9000	0.0000
0.3333	0.3333	0.0500	0.1000	0.9000	0.0146
0.3333	0.3333	0.1000	0.0500	0.9000	0.0166
0.3333	0.3333	0.1000	0.1000	0.9500	0.0169
0.3333	0.3333	0.1000	0.0500	0.9500	0.0205
0.3333	0.3333	0.0500	0.0500	0.9000	0.0225
0.3333	0.3333	0.0500	0.1000	0.9500	0.0238
0.3333	0.3333	0.0500	0.0500	0.9500	0.0268
0.2654	**0.0132**	**0.1000**	**0.4000**	**0.6000**	**0.0275**
0.2654	**0.0132**	**0.0500**	**0.4000**	**0.6000**	**0.0293**

Example: Uniqueness of Target Parameters

- Test case: $P_1 = P_2 = 1/3$; $(\pi_1,\pi_2,\pi_3) = (.1,.1,.9)$

- **Unconstrained** (wider) Grid with 10 expected gradients closest to 0:

P_1	P_2	π_1	π_2	π_3	Norm
0.3333	0.3333	0.1000	0.1000	0.9000	0.0000
0.2119	**0.5761**	**0.1000**	**0.5000**	**0.3000**	**0.0268 (#)**
0.2119	**0.5761**	**0.1000**	**0.5000**	**0.1000**	**0.0305**
0.0132	*0.2654*	*0.1000*	*0.9000*	*0.3000*	*0.0361*
0.2119	**0.5761**	**0.3000**	**0.5000**	**0.1000**	**0.0378**
0.0132	*0.2654*	*0.5000*	*0.1000*	*0.5000*	*0.0382*
0.0132	*0.2654*	*0.3000*	*0.9000*	*0.3000*	*0.0385*
0.0132	*0.2654*	*0.3000*	*0.1000*	*0.5000*	*0.0387*
0.1554	***0.4223***	***0.1000***	***0.7000***	***0.3000***	***0.0395***
0.0132	*0.2654*	*0.7000*	*0.1000*	*0.5000*	*0.0403*

> Frame of reference: norm(0.3333,0.3333,0.05,0.05,0.95) = 0.0268 (#)

Example: Uniqueness of Target Parameters

- Actual basic functioning data

- **Constrained** Grid (P,π) with 10 expected gradients closest to 0:

P_1	P_2	π_1	π_2	π_3	Norm
0.4223	0.4223	0.0500	0.3000	0.9000	0.0170
0.4223	**0.2900**	**0.0500**	**0.1000**	**0.8000**	**0.0177**
0.4223	0.4223	0.0500	0.3000	0.8000	0.0183
0.4223	0.4223	0.1000	0.3000	0.9000	0.0198
0.4223	0.4223	0.0500	0.3000	0.7000	0.0209
0.4223	**0.2900**	**0.1000**	**0.1000**	**0.8000**	**0.0209**
0.4223	0.4223	0.1000	0.3000	0.8000	0.0228
0.4223	0.4223	0.1000	0.3000	0.9500	0.0235
0.4223	0.4223	0.0500	0.3000	0.9500	0.0246
0.4223	0.4223	0.1000	0.3000	0.7000	0.0248

DISCUSSION

- 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
- What's next?
  - — Uniqueness of target: Displays, complicated models
  - — Implications: Delineation of plausible models

DISCUSSION

- A primary issue: Why a hierarchical model at all?
  — e.g. PTSD: Why not DSM *Y*, delineate its 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. In fact: LCR and DSM about equally subject to NDM
     c. Ultimate recommendation: DSM

- Some other issues
  — A seduction: Accuracy property re V_i only for model fit in first stage
  — Why not be Bayesian?
  — Should one be parsimonious or complex?