Hybrid Latent Class Measurement of Health States Lacking a Gold Standard

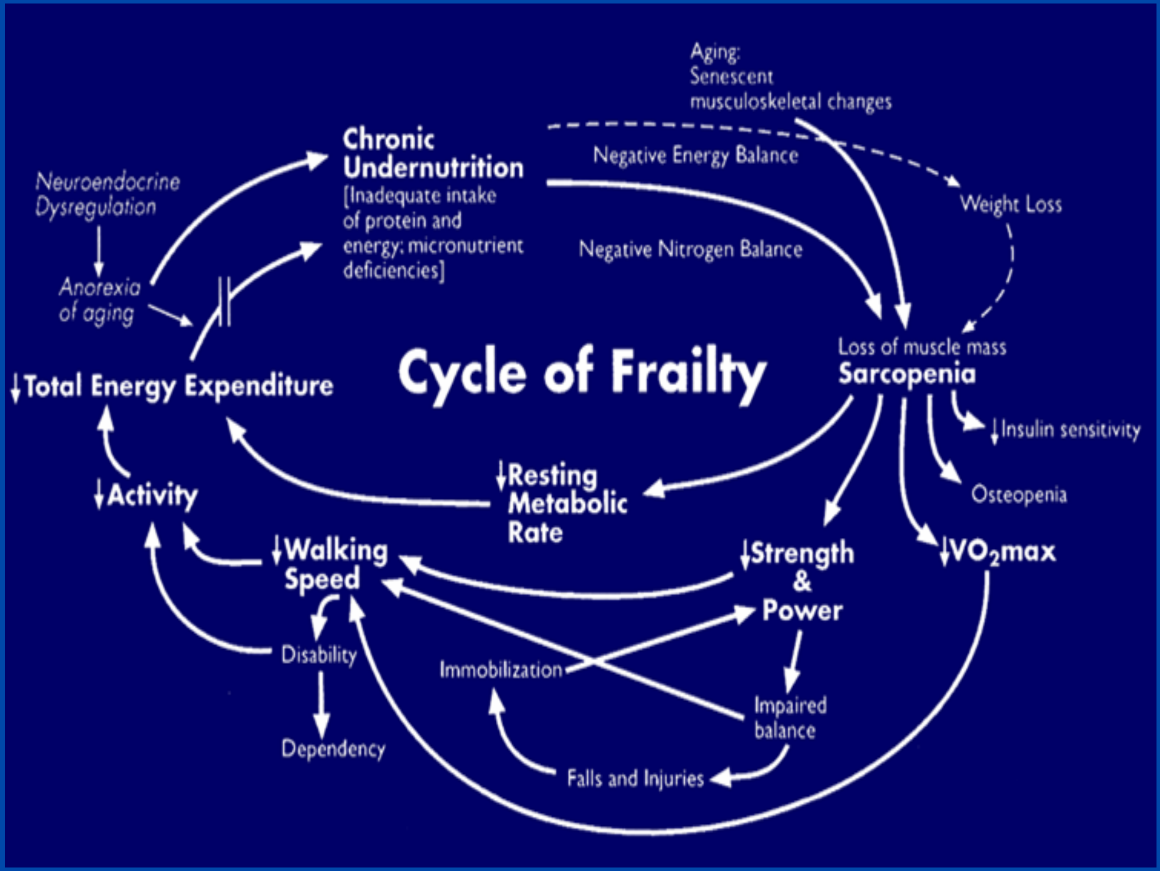
Karen Bandeen-Roche[1]
Jeannie-Marie Sheppard[2]
Jing Ning[1]
Departments of Biostatistics[1] and Mental Health[2]
Johns Hopkins Bloomberg School of Public Health

ENAR Spring Meeting
Atlanta, Georgia
March 13, 2007

Measurement problem

- A health state
  - recognizable
  - conceptually well defined
  - has known consequences
- No gold standard
  - more than diagnostic error
  - no single consensus measurement
  - multifaceted consequences

Measurement Problem
Geriatric Frailty



Fried et al., J Gerontol 2001; Bandeen-Roche et al., J Gerontol, 2006

Measurement Problem
Aging

- Recognition
  - Chronic disease, disability, events
  - Variability among individuals
- Theory: a biological process
  - More than consequence accumulation
  - Multisystem dysregulation
- No gold standard
  - Even to the point of “surrogates”

Successful measures
Classical Approach: Validity

- Face: recognition
- Content: facets covered
- Criterion: utility
- Construct: theory
  - internal; external

DeVellis, 1991; Bartholomew, 1996

Points of the Introduction

- A well defined target; a less-well-defined operationalization

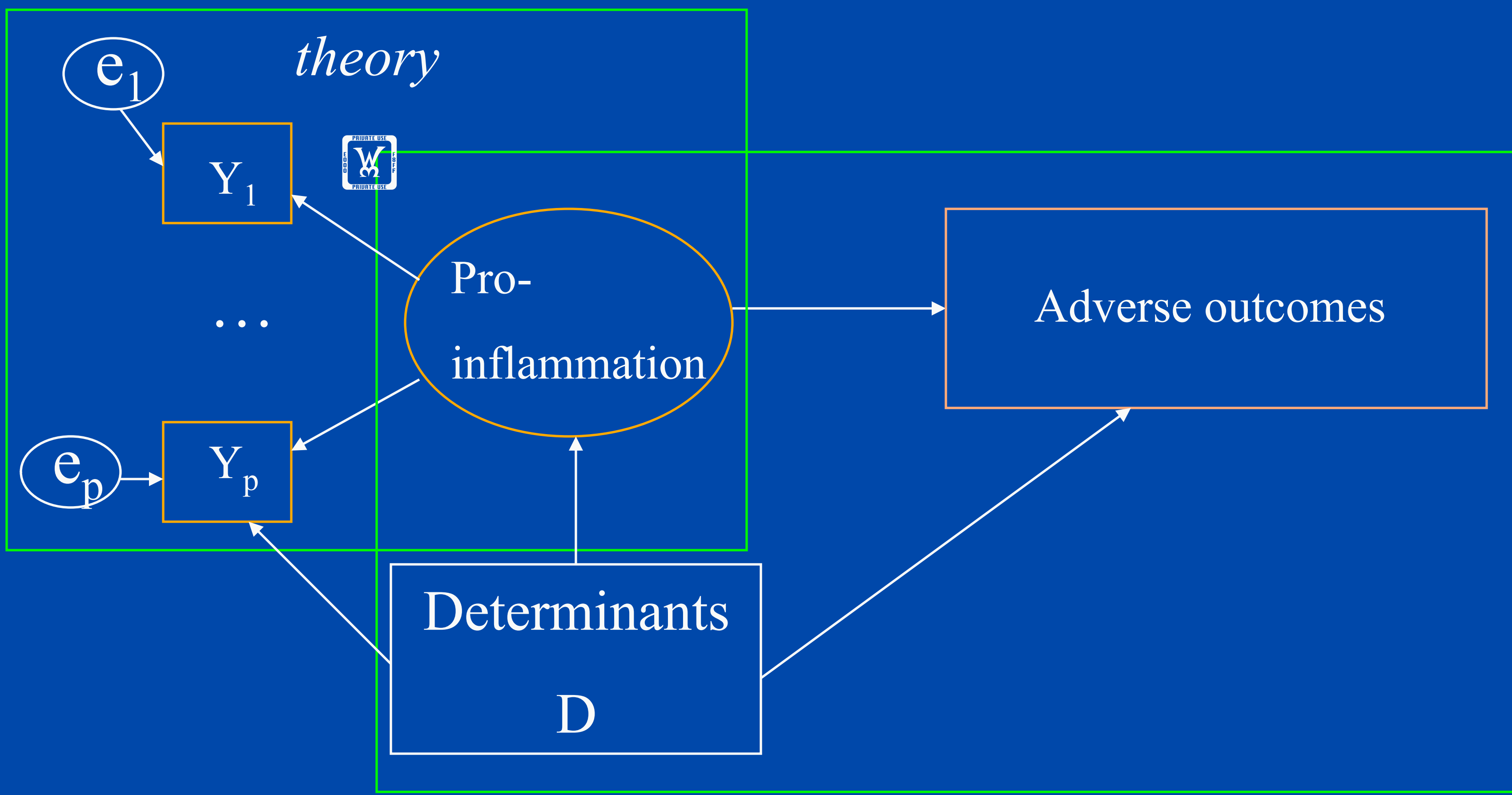
- Will retain validation for measure definition; performance evaluation

- Objective: Method to unify multiple validation aspects in 1 analysis

Outline

- Latent variable paradigm for measurement
- A new idea
  - Aims to balancing potentially conflicting validation premises
  - Application
- Discussion

Measurement
Latent Variable Paradigm



Model

! **Generic**

Kernel

Mixing

! **Specific** (Latent Class Reg.; Categorical U=j , {1,...,J})

! **Measurement assumptions** : $[Y_i|U_i,x_i]$

— conditional independence, nondifferential measurement

> ***heterogeneity in criterion presentation unrelated to measured or unmeasured characteristics***

> ***fundamentally identifying***

Latent Class Measurement
How to obtain "indices"?

- Via **posterior probabilities** of class membership =

$$\hat{F}_{U|Y,x}(u \mid y, x)$$

- Then: exactly how?
  - "Modal": by highest probability
  - "Pseudo-classes": Randomize (*Bandeen-Roche et al., 1997; Wang et al., 2005)*

In what sense is LCA a "measurement" model?

- ~~Does it "discover" structure?~~

- It operationalizes theory
  - Science: Test if predictions borne out
  - Most frequent theory: Homogeneity

Latent Class Measurement
Syndrome Validation Application

- Criteria manifestation is syndromic

  "a group of signs and symptoms that occur together and characterize a particular abnormality" *(Webster Medical Dictionary 2003)*

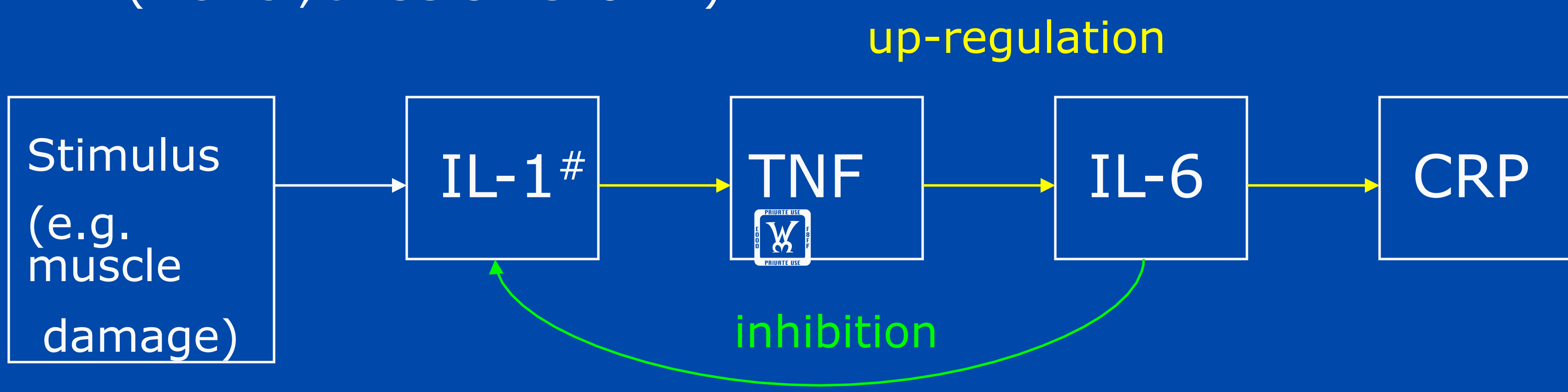
- If criteria characterize syndrome:
  - At least two clinically homogeneous groups (if <2, no co-occurrence)
  - No subgrouping of symptoms (otherwise, more than one abnormality characterized)

Bandeen-Roche et al., J. Gerontol Med Sci, 2006

Measurement Application: Pro-Inflammation

- Central role: cellular repair

- A hypothesis: dysregulation key in adverse aging
  - Muscle wasting (*Ferrucci et al., JAGS 50:1947-54; Cappola et al, J Clin Endocrinol Metab 88:2019-25*)
  - Receptor inhibition: erythropoetin production / anemia (*Ershler, JAGS 51:S18-21*)



\# Difficult to measure. IL-1RA = proxy

Rationale of the New Work

- Which deserves pre-eminence?
  - Internally validating assumptions?
  - Externally validating assumptions?
    - Frailty: close tie to systemic dysregulation
    - Depression: genetic "subtypes"
    - Aging: tie to chronological age
  - Some compromise?

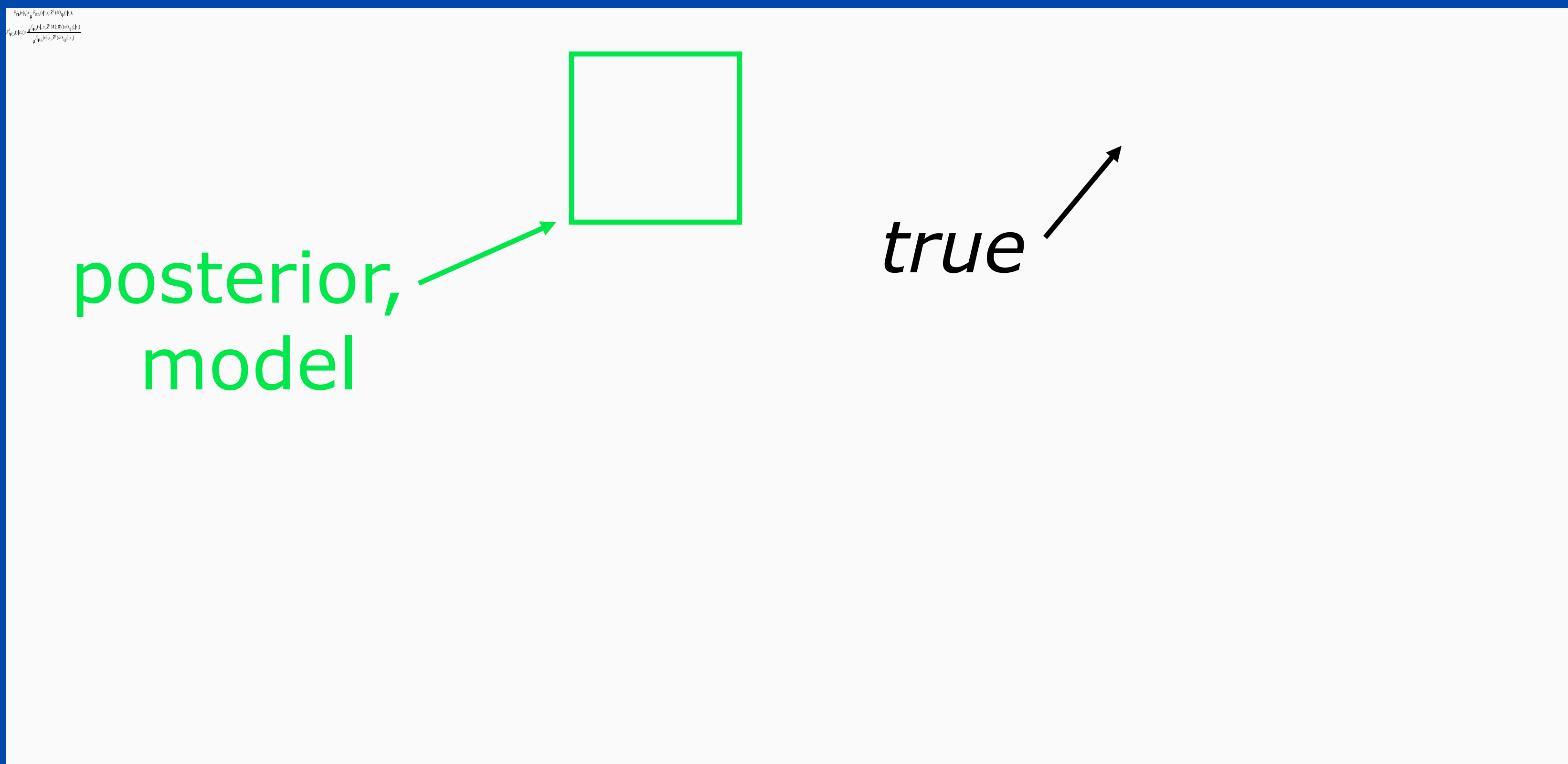
Rationale of the New Work

- Which deserves pre-eminence?
  - Internally validating assumptions
  - Externally validating assumptions?
  - Some compromise?

- A model (LCR) including externally validating variables and fitting by ML already "is" a compromise

A representation theorem

- Consider “mixing” & “kernel” distributions:



A representation theorem

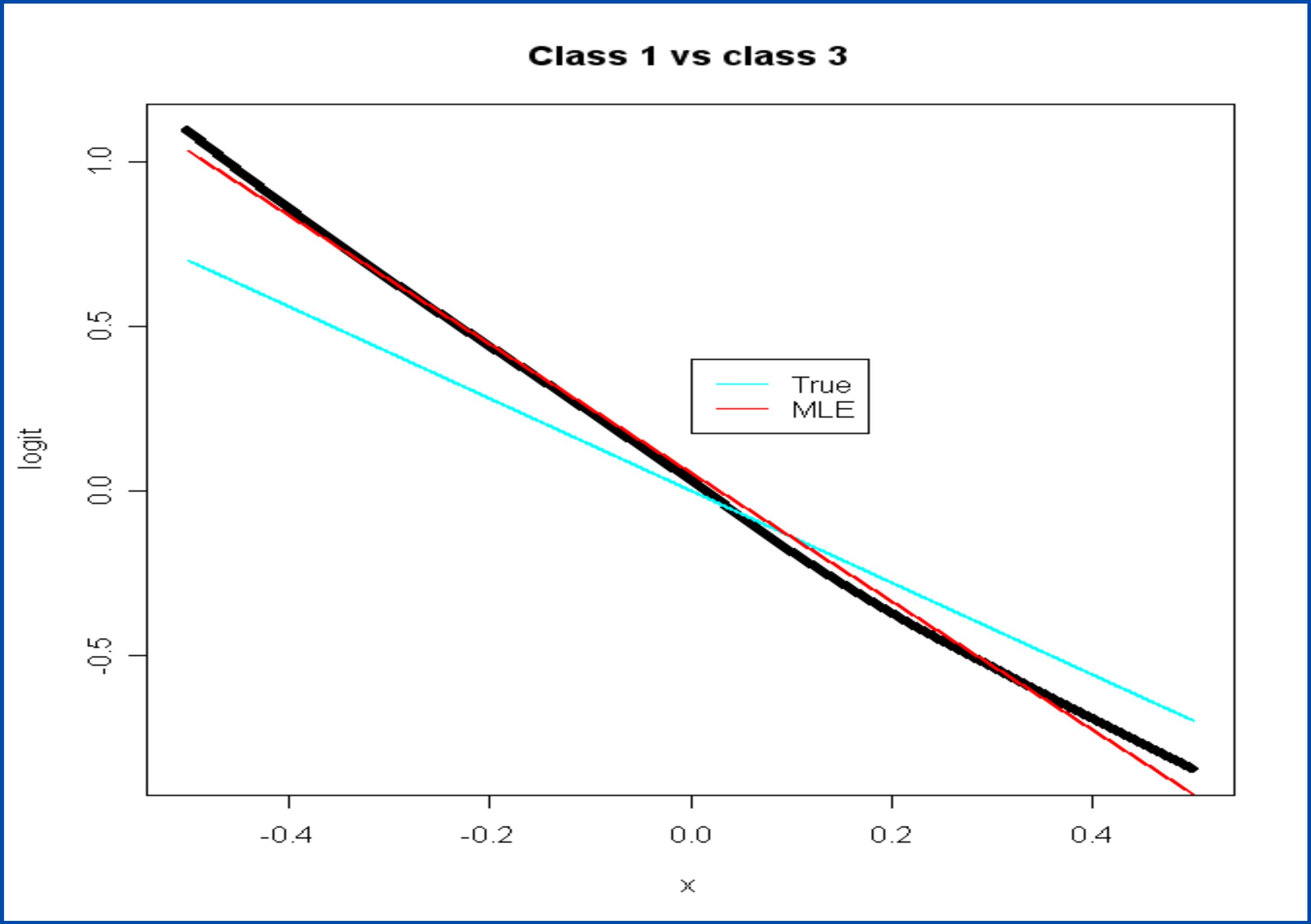
- Y_i is equivalent in distribution to Y* constructed as

1) Generate V_i* from $F^*_{V|x}(v \mid x_i)$

2) Given V_i*, generate Y* from $F^*_{Y|V,x}(y \mid V_i^*, x_i)$

- Relevance:
  - True for θ^* = Huber (1967) limit of MLE (e.g.)

True vs. realized mixing models



Rationale of the New Work

- Which deserves pre-eminence?
  - Internally validating assumptions
  - Externally validating assumptions?
  - Some compromise?

- Proposal: Allow stronger (or weaker) compromise than ML via "penalized" fitting

Implementing penalization

- On LCR kernel: Houseman, Coull & Betensky, *BMCS* online early
- On LCR mixing distribution: Sheppard Ph.D. thesis
- Key questions
  - Form of the penalty
  - Different purpose than usual?
  - What is the objective function?

Penalization
Very brief background

- Fitting: minimize

$$-2 \ln L(\theta;Y,x) + \lambda g(\theta)$$

- Examples
  - "Ridge": $g(\theta) = \Sigma_j \theta_j^2$
  - "Lasso": $g(\theta) = \Sigma_j |\theta_j|$

Green, Int Stat Rev, 1987; Tibshirani, JRSS-B, 1996

Penalization
Very brief background

- A useful equivalence: penalized fit obtains via formulating parameters as crossed random effects
  - "Ridge": $\theta_j \sim N(0,\sigma\{\lambda\}^2)$
  - "Lasso": $\theta_j \sim \text{double exp}(0,h\{\lambda\})$

Wahba, JRSS-B, 1978; Ngo & Wand, J Stat Software, 2004

Form of the penalty
Current case

- Usual purpose: regularization

- Here: secondary validation

- Discriminant hypothesis: Genotypes predispose individuals to only one “subtype” of depression

Form of the penalty
Genetic subtypes example

- Say, LCR with one normal class (1) and two disordered classes (2, 3):

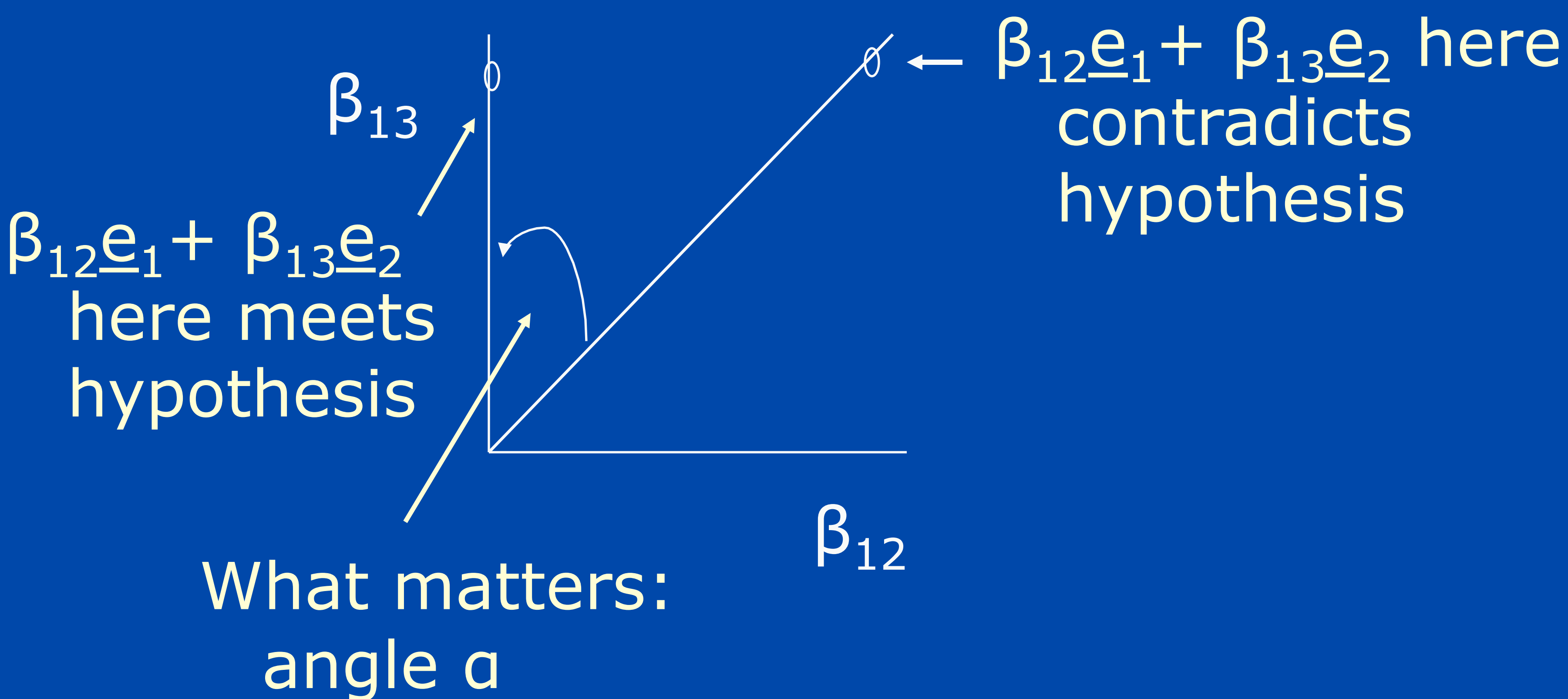
- Hypothesis: β_{1j} negligible, and $\beta_{1j'}$ appreciable, in

$$\log\left(\frac{p_k}{p_1}\right) = \beta_{0k} + \beta_{1k}x$$

with p_k = pr(class k); x=genotype indicator; k=2,3; j, j′ ϵ {2,3}; j ≠ j′

Form of the penalty
Genetic subtypes example

- Ridge, lasso not quite right

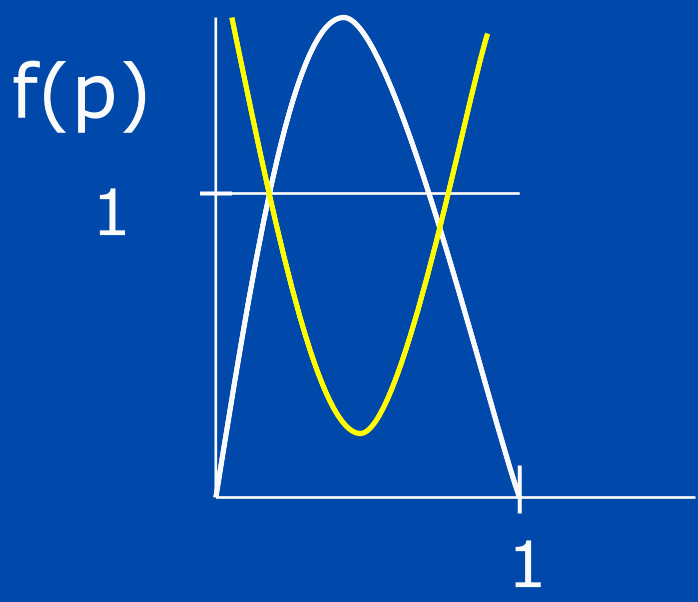


Form of the penalty
Genetic subtypes example

- Approach 1
  - Consider α ϵ [0,90] degrees
  - Desired orientations are cos(α)=1, sin(α)=1
  - i.e., goal: minimize cos(α)+sin(α)
  - i.e. minimize $$\frac{|\beta_{12}|+|\beta_{13}|}{\sqrt{\beta_{12}^2+\beta_{13}^2}}$$

Form of the penalty
Genetic subtypes example

- Approach 2
  - Write $\beta_{12} = p\beta$; $\beta_{13} = (1-p)\beta$
  - Fit with beta random effect on p



  - Generalization: $\underline{\beta} = \underline{p}\beta$, $\underline{p}$ ~ Dirichlet

Fitting
Approach 2

- E-M algorithm: quite straightforward

- E-step: Computes posterior class membership probabilities given current parameter iterates

- M-step: minimize (e.g. Nelder-Mead)

$$-\sum_{i=1}^{n}\sum_{j=1}^{J} h(j|data)\ln[f_{U|x}(u|x,p,\beta)] + (1-\frac{\Delta}{2})\ln[p(1-p)]$$

Simulation study
Three-class model

- 100 reps; single x~Unif(-.5,.5); n=1000
- Poly Log Reg: $\beta_{01}=\beta_{02}=0$; $\beta_{13}=-1.4$;
  $\beta_{12}\epsilon\{0,-0.5,-1.4\}$
- Measurement:

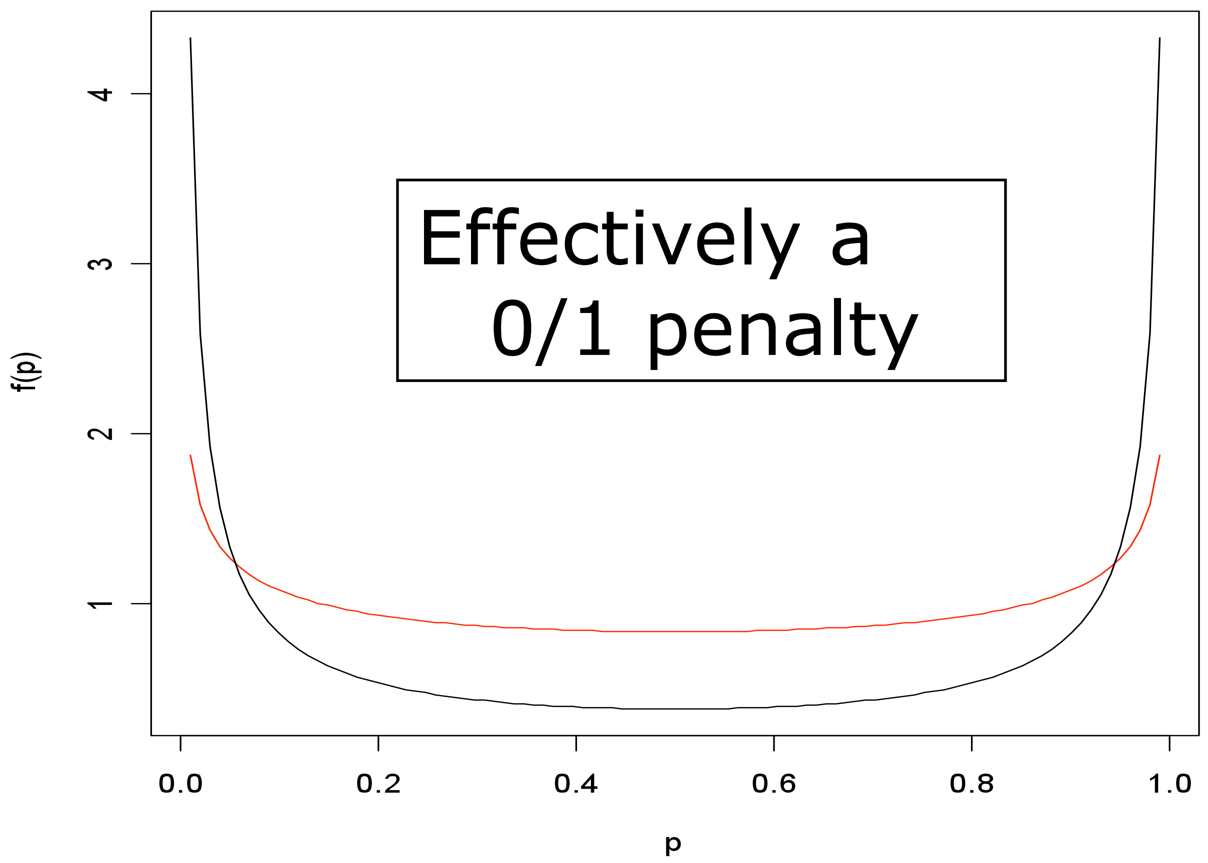
Class 1	Class 2	Class 3
.15	.85	.85
.15	.85	.85
.15	.85	.85
.15	.13	.85
.15	.13	.85

Simulation study
Three-class model

- Two assumption scenarios
  - Frank LCR
  - Differential measurement: First three items have increased log(odds =1) per unit x of 1.4 within each class

Simulation study

Beta model: Δ = 1.5, .5



Simulation Study
Diff. Meas.; β_{12}=0; β_{13}=-1.4

Param.	Penalized		LCR	
	Estimate	SE	Estimate	SE
β_{12}	-0.04	0.14	-0.54	0.31
β_{13}	-0.79	0.30	-1.01	0.34

Simulation Study
Non-diff meas; β_{12}=0; β_{13}=-1.4

Param.	Penalized		LCR	
	Estimate	SE	Estimate	SE
β_{12}	0	0	0.04	0.32
β_{13}	-1.42	0.35	-1.41	0.38

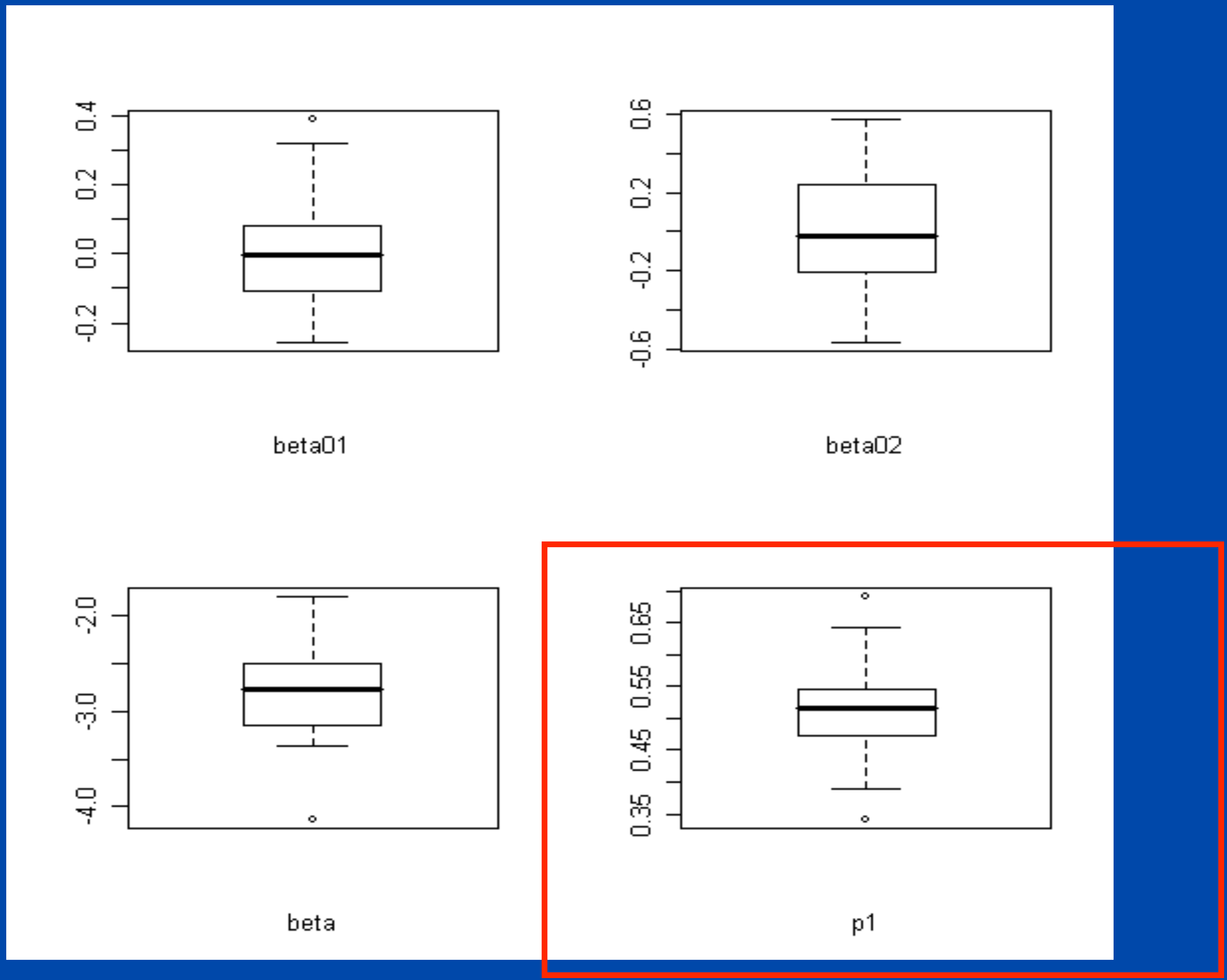
Simulation Study
Diff. Meas.; $\beta_{12}=\beta_{13}=-1.4$

Param.	Penalized		LCR	
	Estimate	SE	Estimate	SE
β_{12}	-1.61	0.32	-2.00	0.31
β_{13}	-0.08	0.28	-1.02	0.34

Simulation Study
Non-diff meas; $\beta_{12}=\beta_{13}=-1.4$

Param.	Penalized		LCR	
	Estimate	SE	Estimate	SE
β_{12}	-1.45	0.34	-1.45	0.30
β_{13}	-1.38	0.31	-1.38	0.31

Simulation Study
Non-diff meas; $\beta_{12}=\beta_{13}=-1.4$



One empirical lead

Deciding the extent of penalization

- Notice the form of $F^*_{V|x}(v \mid x_i)$:

$$f^*_{V|x}(v|x_i) = f_{V|x}(v|x_i) \int \frac{f_{\theta|V,x}(\theta|v,x_i;\lambda^*)}{f_{\theta|x}(\theta|x_i;\lambda^*)} dG_{\theta|x}(\theta|x_i).$$

Factor by which f^*, f differ at x_i

- Idea 1: Right penalty yields $f^* = f$

Simulation study
Three-class model

- Small: 100 reps; single x~Unif(-.5,.5)
- Multiple n: Here, =2000
- Poly Log Reg: $\beta_{01}=\beta_{02}=0$; $\beta_{12}=-1.4$; $\beta_{12}=-2.8$
- Measurement:

Class 1	Class 2	Class 3
.15	.85	.85
.15	.85	.85
.15	.85	.85
.15	.13	.85
.15	.13	.85

Simulation study
Three-class model

- Two scenarios (among more)
  - Frank LCR
  - Differential measurement: last two items have increased log(odds =1) per unit x of 1.4 within each class
- Premise: $f^{*}_{V|x}(v \mid x_i, \theta)$, $f_{V|x}(v \mid x_i, \theta)$ quite different

- Measure: Kullback-Leibler distance

KL Distance: f*, f

Scenario 1, n=2000

$\hat{\beta}_{12}$ \ $\hat{\beta}_{22}$	-3.4	-3.3	-3.2	-3.1	-3.0	-2.9	-2.8	-2.7	-2.6	-2.5	-2.4	-2.3	-2.2
-2.0	4.99	4.76	4.76	4.86	4.89	5.15	5.26	5.42	6.23	6.34	6.93	7.59	7.99
-1.9	4.58	4.28	4.40	4.57	4.19	4.42	4.62	5.09	5.15	5.62	6.03	6.91	7.31
-1.8	4.52	4.36	4.18	4.07	3.88	3.96	4.22	4.26	4.55	5.09	5.52	5.96	6.58
-1.7	4.30	4.05	3.90	3.64	3.85	3.71	3.73	4.05	4.35	4.46	4.92	5.33	5.77
-1.6	4.56	4.21	3.80	3.62	3.52	3.54	3.67	3.69	3.88	4.07	4.36	4.88	5.46
-1.5	4.67	4.11	3.88	3.70	3.56	3.41	3.46	3.42	3.75	3.74	4.28	4.52	4.85
-1.4	4.87	4.39	3.91	3.84	3.62	3.27	3.62	3.40	3.69	3.68	3.70	4.03	4.52
-1.3	5.25	4.73	4.50	4.16	3.86	3.54	3.45	3.46	3.39	3.52	3.78	4.12	4.43
-1.2	5.58	4.99	4.76	4.47	4.16	3.81	3.70	3.60	3.75	3.74	3.85	4.25	4.30
-1.1	6.25	6.05	5.26	4.90	4.55	4.14	4.20	4.03	4.01	3.94	3.91	4.45	4.28

True

ML

KL Distance: f*, f

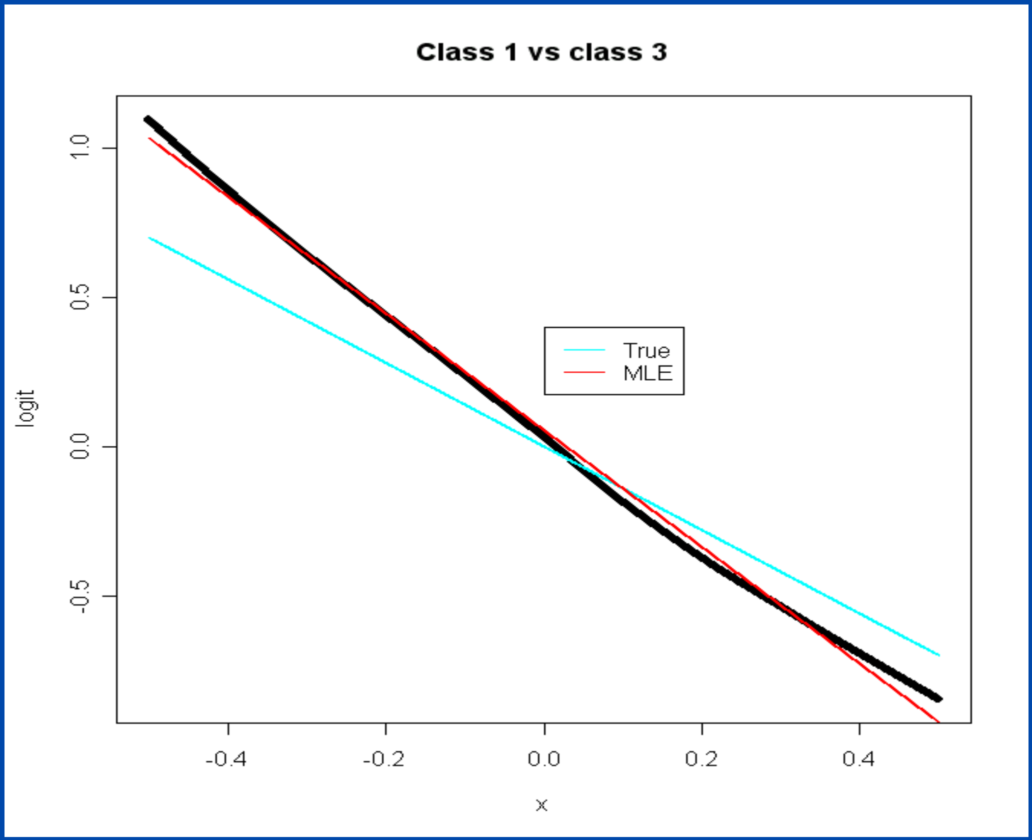
Scenario 2, n=2000

$\hat{\beta}_{12}$ \ $\hat{\beta}_{22}$	-3.8	-3.7	-3.6	-3.5	-3.4	-3.3	-3.2	-3.1	-3.0	-2.9	-2.8	-2.7	-2.6
-2.4	4.03	4.37	4.63	5.05	5.39	5.93	6.35	7.17	8.00	8.76	9.36	10.40	11.74
-2.3	3.79	3.87	4.10	4.59	4.93	5.14	5.84	6.38	6.76	7.79	8.55	9.46	10.50
-2.2	3.48	3.63	3.90	3.98	4.27	4.60	5.20	5.76	6.17	7.01	7.78	8.26	9.65
-2.1	3.31	3.17	3.47	3.51	3.95	4.25	4.69	5.04	5.64	6.34	7.01	8.09	9.07
-2.0	3.19	3.29	3.41	3.33	3.70	3.94	4.34	4.60	5.10	5.62	6.70	7.24	8.02
-1.9	3.17	3.09	3.19	3.27	3.39	3.64	3.99	4.25	4.93	5.40	6.17	6.90	7.37
-1.8	3.31	3.24	3.22	3.26	3.35	3.63	3.98	4.35	4.75	5.12	5.34	6.40	7.00
-1.7	3.56	3.33	3.43	3.32	3.31	3.57	3.85	4.17	4.40	4.79	5.43	6.00	6.33
-1.6	3.83	3.77	3.60	3.69	3.68	3.62	3.80	4.19	4.65	4.87	5.38	6.21	6.62
-1.5	4.36	3.95	4.02	3.97	3.89	3.82	4.05	4.24	4.56	5.05	5.37	5.86	6.36
-1.4	4.90	4.69	4.43	4.28	4.34	4.46	4.35	4.65	4.88	5.11	5.41	5.99	6.49
-1.3	5.56	5.41	5.11	4.95	4.77	4.84	4.72	4.74	5.01	5.49	5.85	6.19	6.60
-1.2	6.41	5.97	5.87	5.59	5.37	5.17	5.33	5.18	5.52	5.96	6.08	6.31	6.99

ML

True

Simulation Study
Empirical support for "penalty"?



- Average conditional probability estimates amazingly stable
- Distinction: $Y|V^*,x$

Frailty analysis: Data

InCHIANTI (*Ferrucci et al., JAGS, 48:1618-25*)

- **Aim** : Causes of walking decline
- Brief design
  - Random sample ≥ 65 years (n=1270)
  - Enrichment for oldest-old, younger ages
  - Participation: > 90% in the primary sample
  - Home interview, blood draw, physical exam
- **Dysregulation: inflammation – 7 cytokines**
  - *IL-6, CRP, TNF-α, IL-1RA, IL-18, IL-1B, TGF-β*
  - Here: concern = poorer inhibition
- **Frailty: Fried criteria (as before)**

Frailty analysis: Results

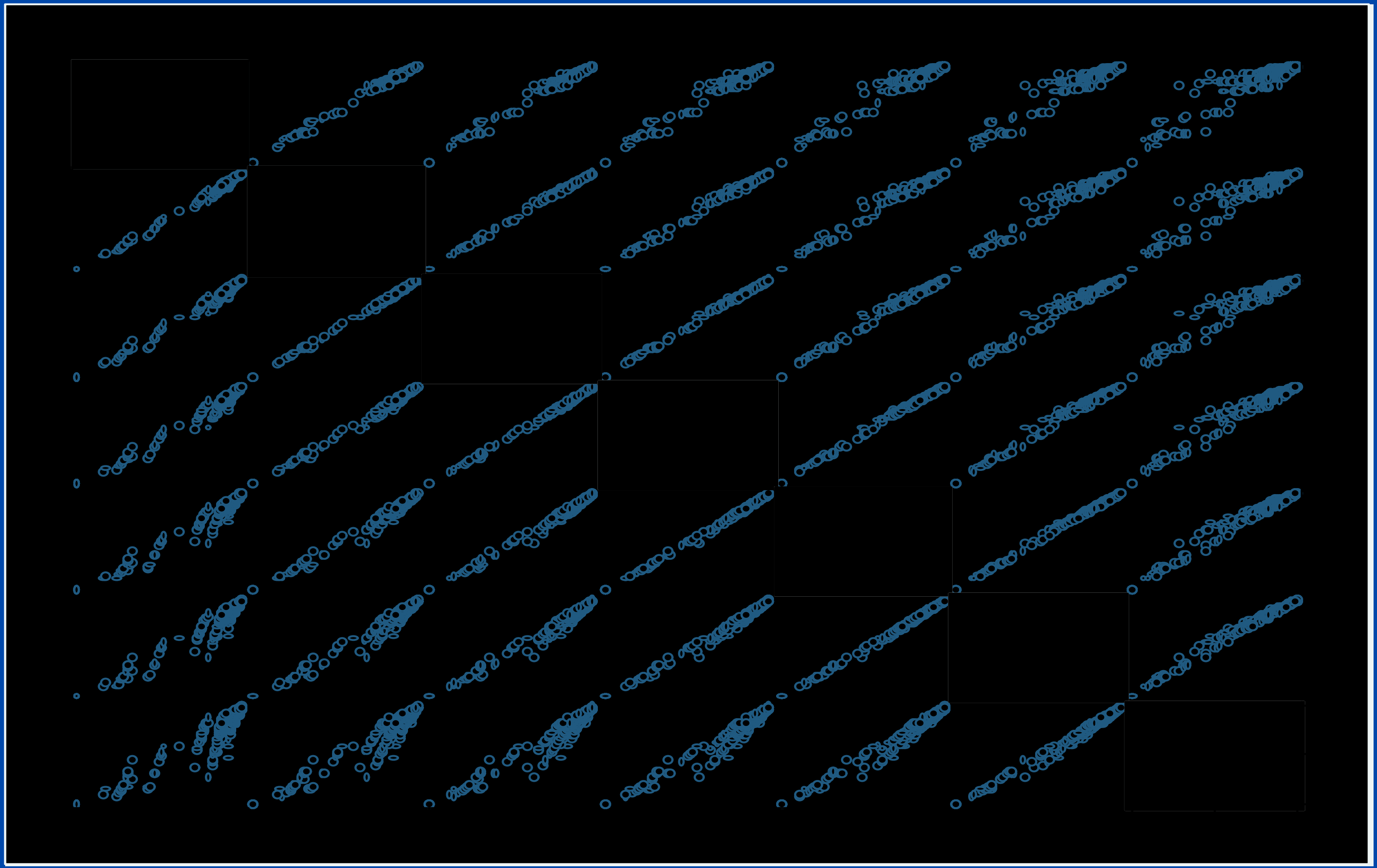
- Measurement model: 2 classes
  - Conditional probabilities similar to WHAS
  - Lower “frail” prevalence (15% vs. 27%)

- Regression model
  - 1 SD worse inhibition index associated with 35% reduction in non-frail odds (z ~ 3)
  - Regression coefficient on original index scale: 3.00

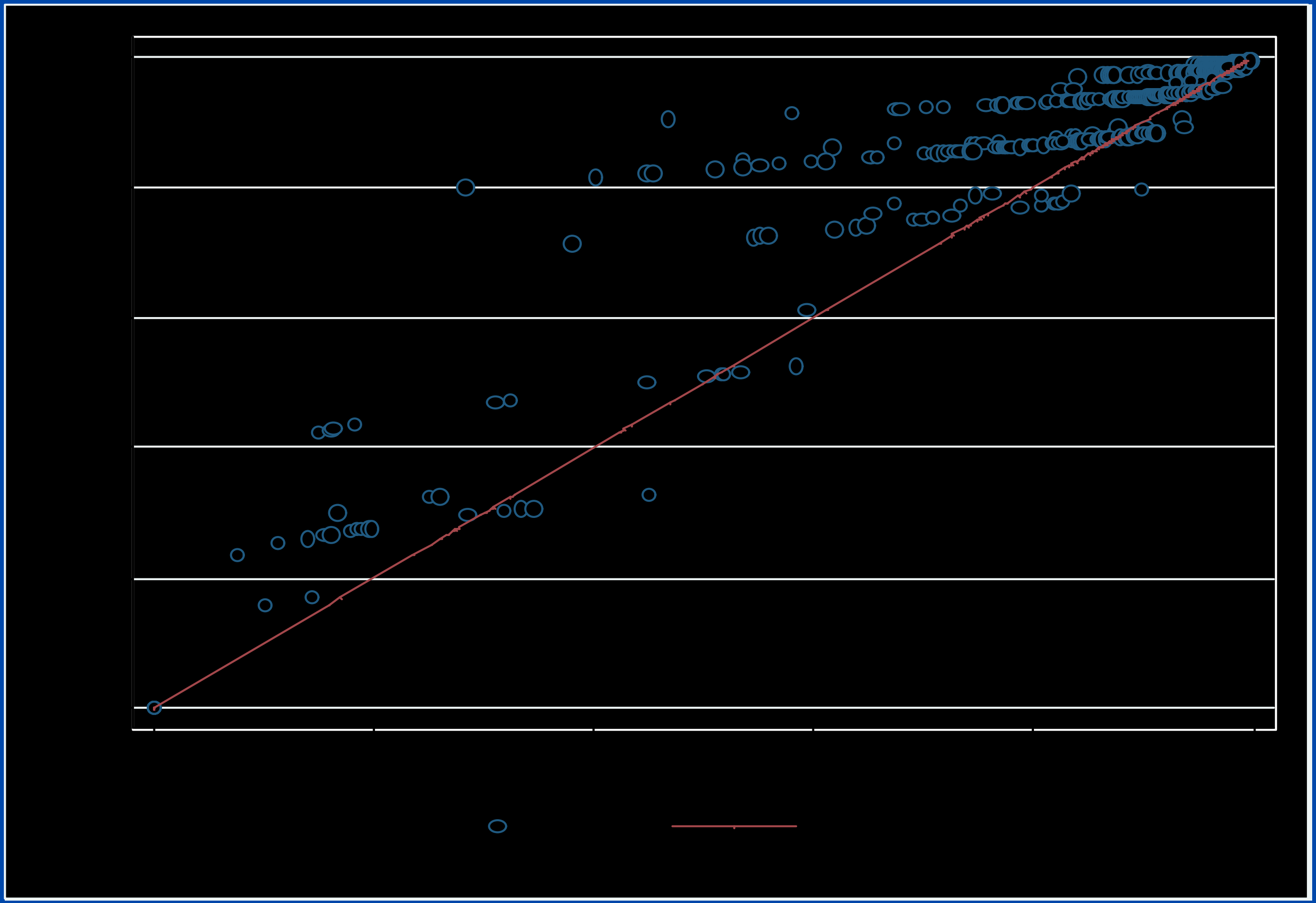
- Next: Vary regression coefficients in increments of +/- 0.5, up to +/- 2.0

Frailty analysis: Results

Posterior probs. from different fits



Frailty analysis: Results
Posterior probs. non-frail, different fits



Frailty analysis: Results

Age-adjusted relation to mobility

Frailty fit: inflam. slope	Mobility slope (frail vs non)	SE
ML – 2.0	-1.1	.089
ML – 1.0	-1.0	.087
ML – 0.5	-1.0	.086
ML	-0.99	.085
ML + 0.5	-0.93	.085
ML + 1.0	-0.92	.085
ML + 2.0	-0.82	.083

Recap

- Presented: Frameworks for measurement
  - of complex geriatric health states
  - incorporating biological knowledge

- Demonstrations
  - Frailty in WHAS
  - Frailty and inflammatory dysregulation in In CHIANTI

Rationale for the proposal

- vs looser internal validation criteria?
  - estimability
- vs Bayesian approach
  - depends on degree of empiricism
  - if balance by “consensus”—Bayesian
- Allows some distrust of the data

Research needed

- Theory elicitation, incorporation
- Methodology freeing measurement model estimation to “move” with “penalty”
  - Rotation?
  - Penalty on conditional probabilities
- Compromise of latent variable, predictive approaches
- Best index derivation

Implications

- Refined understanding of aging states and their measurement
  - Integrating biology
  - Increasing sensitivity, specificity
- Heightened accuracy, precision for
  - Delineating etiology
  - Developing and targeting interventions

Acknowledgments

- Hopkins Colleagues

  Linda Fried, Ron Brookmeyer, Yi Huang, Jeannie-Marie Leoutsakos, Jeremy Walston, Qian-Li Xue, Scott Zeger

- Colleagues outside of Hopkins

  Luigi Ferrucci, Jack Guralnik, Don Ingram, Richard Miller

- Funding / Institutional Support

  Johns Hopkins Older Americans Independence Center, National Institute on Aging, Alliance for Aging Research