On the Study
of Hypothetical Health Constructs
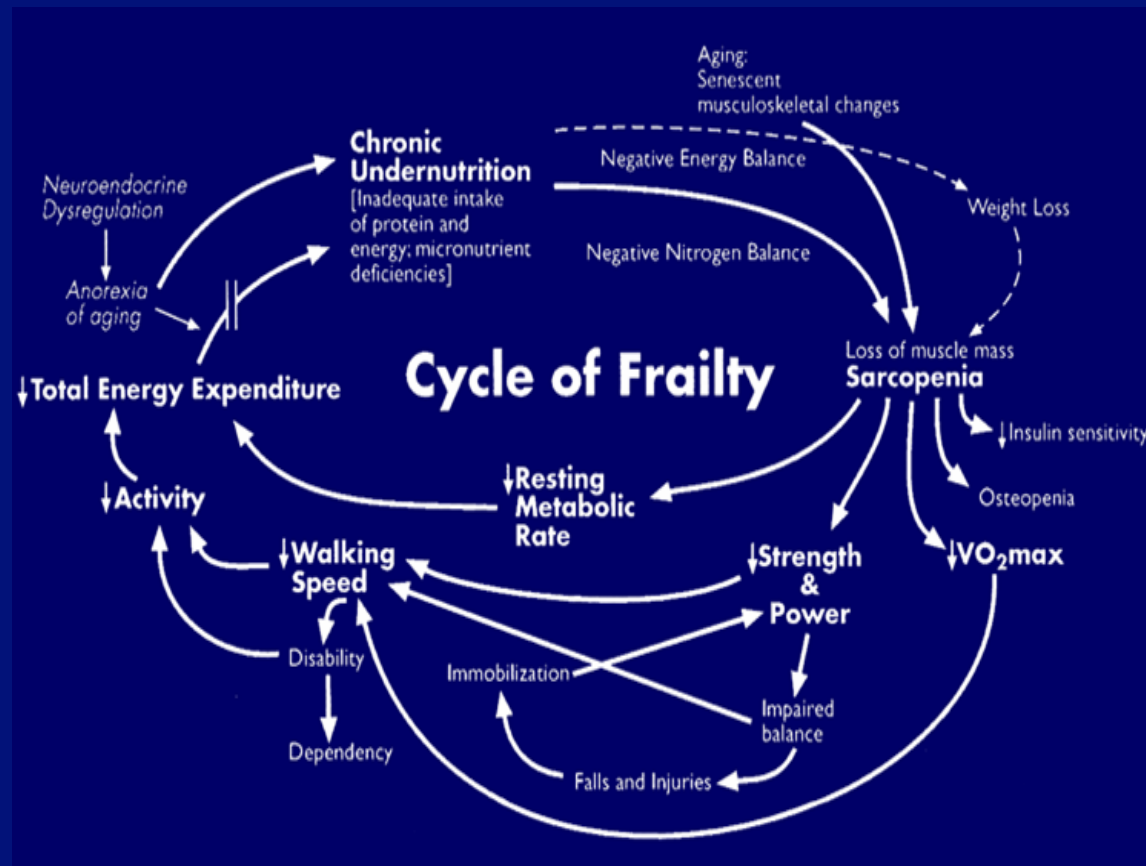
Karen Bandeen-Roche
Department of Biostatistics
Johns Hopkins University

Joint Statistical Meetings
Toronto, Ontario
August 9, 2004

With thanks to: Luigi Ferrucci, Yi Huang

Introduction
Motivation

Fried et al., J Gerontol 56:M146-56, 2001

Ultimate Scientific Aims

- Does frailty exist?
    - More than a marker of disease
    - More than severe disability
    - A *syndrome*: more than the component parts

- Improved measurement
    - Beyond current "criterion count" (*Fried et al. 2001*)
    - Product: a summary variable

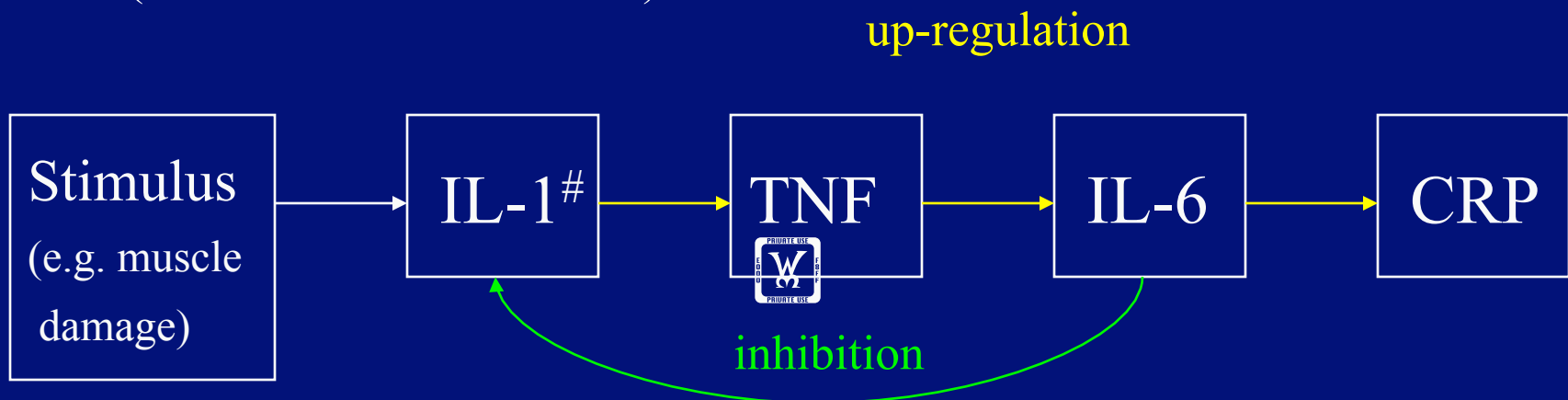# Statistical Contribution to Achievement of Aims

- Long psychometric tradition
  - Validity, (reliability)
- Role of latent variable modeling?
  - Reveal underlying truth?
  - Operationalize theory?
  - Sensitivity analyses?
  - None?
  - Differential measurement

Outline

- Scientific focus: inflammation
  - One component of frailty
- Existence / summary: A paradigm
  - Subject to theory
- Analysis
  - Data: InCHIANTI

Science: Inflammation

- Central role: cellular repair

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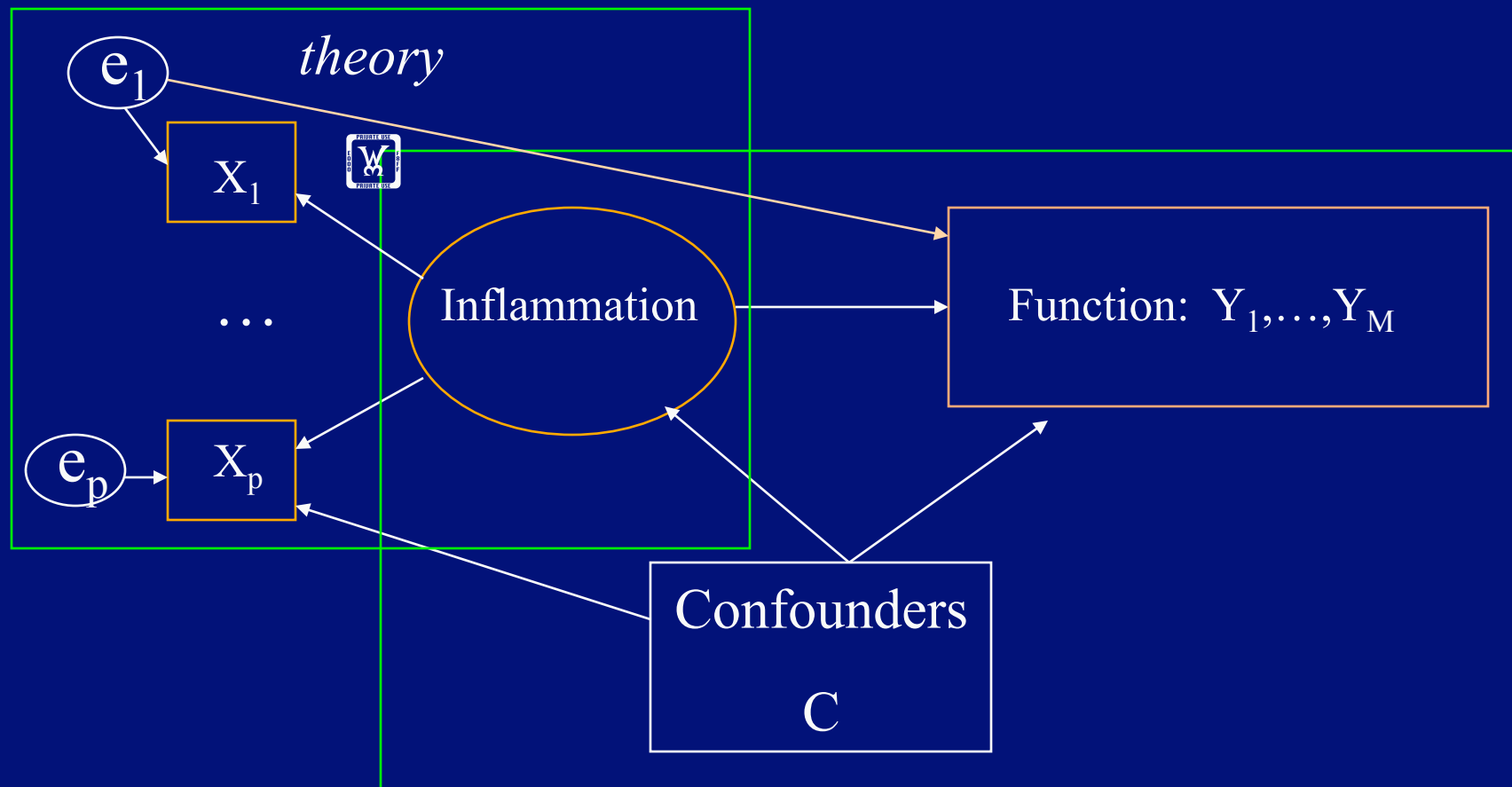 - Receptor inhibition: erythropoetin production / anemia *(Ershler, JAGS 51:S18-21)*

up-regulation

| Stimulus (e.g. muscle damage) | → | IL-1# | → | TNF | → | IL-6 | → | CRP |

inhibition

Difficult to measure. IL-1RA = proxy

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

- **Inflammation – 7 cytokines (5+2)**

  IL-6, CRP, TNF-α, IL-1RA, IL-18,

  (IL-1β, TGF-β)

- **Functional elements – solo; Z-score average**

  Usual & rapid speed; muscle power; range of motion; neurological intactness

- **Confounders**

  Age, gender, (history of: cancer, cardiovascular disease, diabetes, smoking)
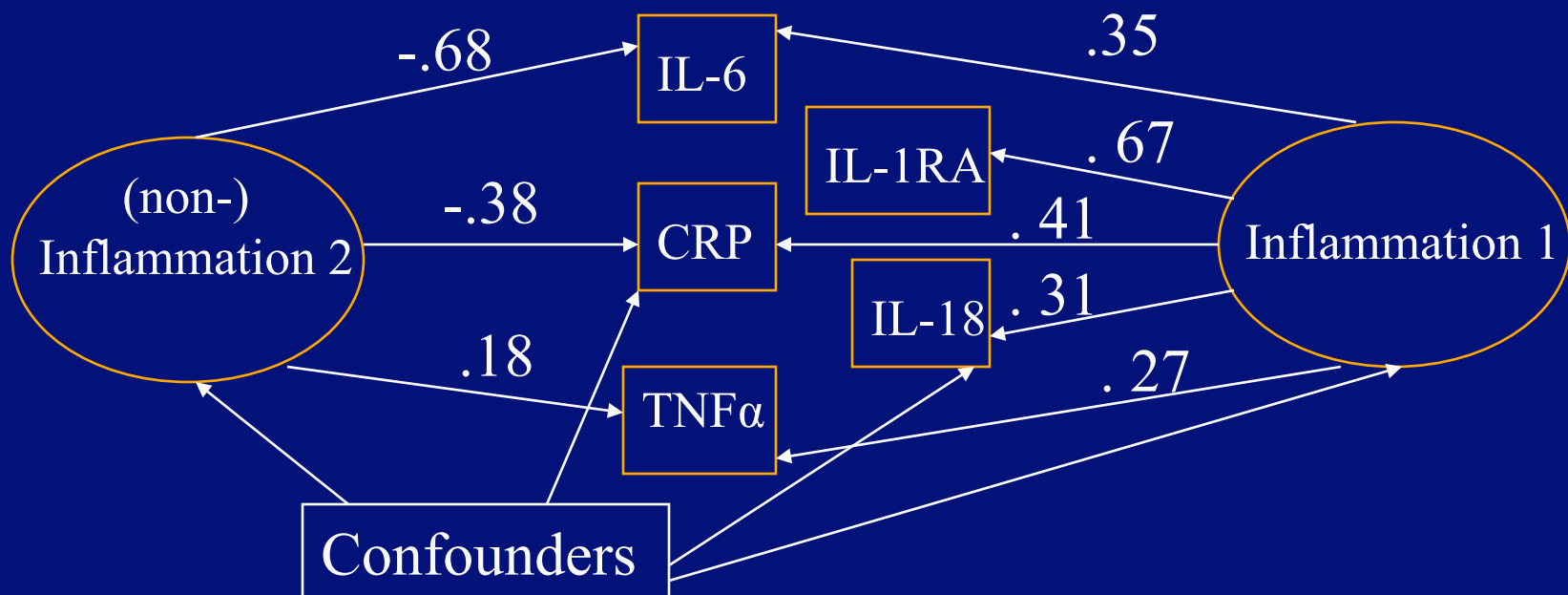
Existence / Summary Paradigm

Statistical Methodology
Construct Definition

- LV method: measured = physiology + noise
  - Multivariate normal underlying variables, errors
  - Conditional independence of errors
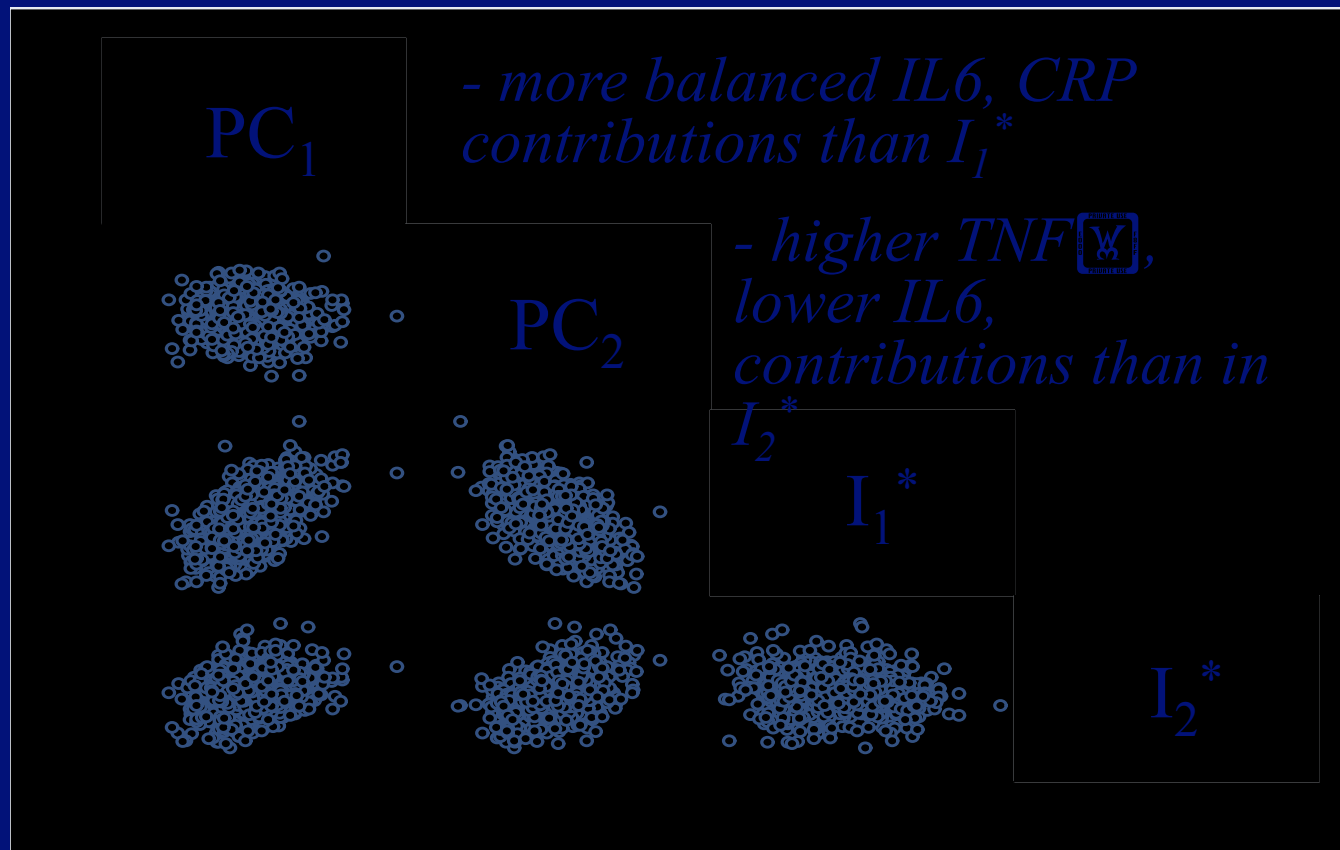
Statistical Methodology
Regression of Functioning on Inflammation

- <u>Method 1</u>: Full LV model (a.k.a. two slides ago)

- <u>Method 2</u> (two-stage; empirical):
  - "Inflammation" values via principal components
  - Regression of functioning on "inflammation"

- <u>Method 3</u> (three-stage; compromise):
  - Fit LV measurement model
  - Random generation of "inflammation" from $[I^*|X,c]$
  - Regression of functioning on "inflammation"

Statistical Methodology
Method 2: Construct Definition

- Principal components

Statistical Methodology
Method 3: Properties (*Bandeen-Roche 2003*)

- Randomization imposes limiting hierarchical model

- $[X|I^*,c]$ arbitrarily well approximates that model

- $[I_j^*|c]$ arbitrarily well approximates $[I_j|c]$

- Implication: Rigorous validation

Findings
Mobility association with inflammation

Function Measure	Analytic method					
	Full LV (*1*)		Empirical (*2*)		Compromise (*3*)	
	I_1	I_2	PC_1	PC_2	I_1^*	I_2^*
Summary	**-.11**	*.07*	**-.10**	-.01	**-.05**	**.07**
Motion	**-.15**	.03	**-.10**	.04	**-.09**	*.06*
Speed	**-.12**	*.08*	**-.11**	-.02	-.02	**.08**
Strength	-.04	-.03	.01	-.03	<-.01	<-.01
Neuro.	-.07	**.11**	**-.30**	-*.12*	**-.14**	*.11*
Note: R^2 f or Methods 2, 3 almost identical						

Cytokine effects
Differential Measurement

- Method 1: Within LV model

- Method 2: PCs + residuals of X on PCs
    - A mess

- Method 3: I^* + residuals of X on I^*

Cytokine effects
Differential Measurement

- Effects & directionality at ∀=.05 level:

Function Measure	Cytokine									
	IL-1RA		IL-18		TNF ∀		IL-6		CRP	
	LV	I*	LV	I*	LV	I*	LV	I*	LV	I*
Summary		.·	.·							
Motion		.·			[	[				
Speed		.·				.·		.·		
Strength									[	[
Neuro.			.·	.·	.·	.·				

Discussion

- How to best use the I*s (pseudo-values)?
    - Randomized versus posterior mode?
    - Validation step only?
    - Measurement error correction?

- Why the differences between "full" and "compromise" approaches?
    - Issues related to previous bullet?

Discussion

- Why the differences between "full" and "compromise" approaches?
  - Issues related to previous bullet?
  - Identification issues: "construct" vs. "error"?
  - "Scoring" anomaly (covariate-based imputation)?
- Definitely needed: an empirical summary

- An opportunity for statisticians