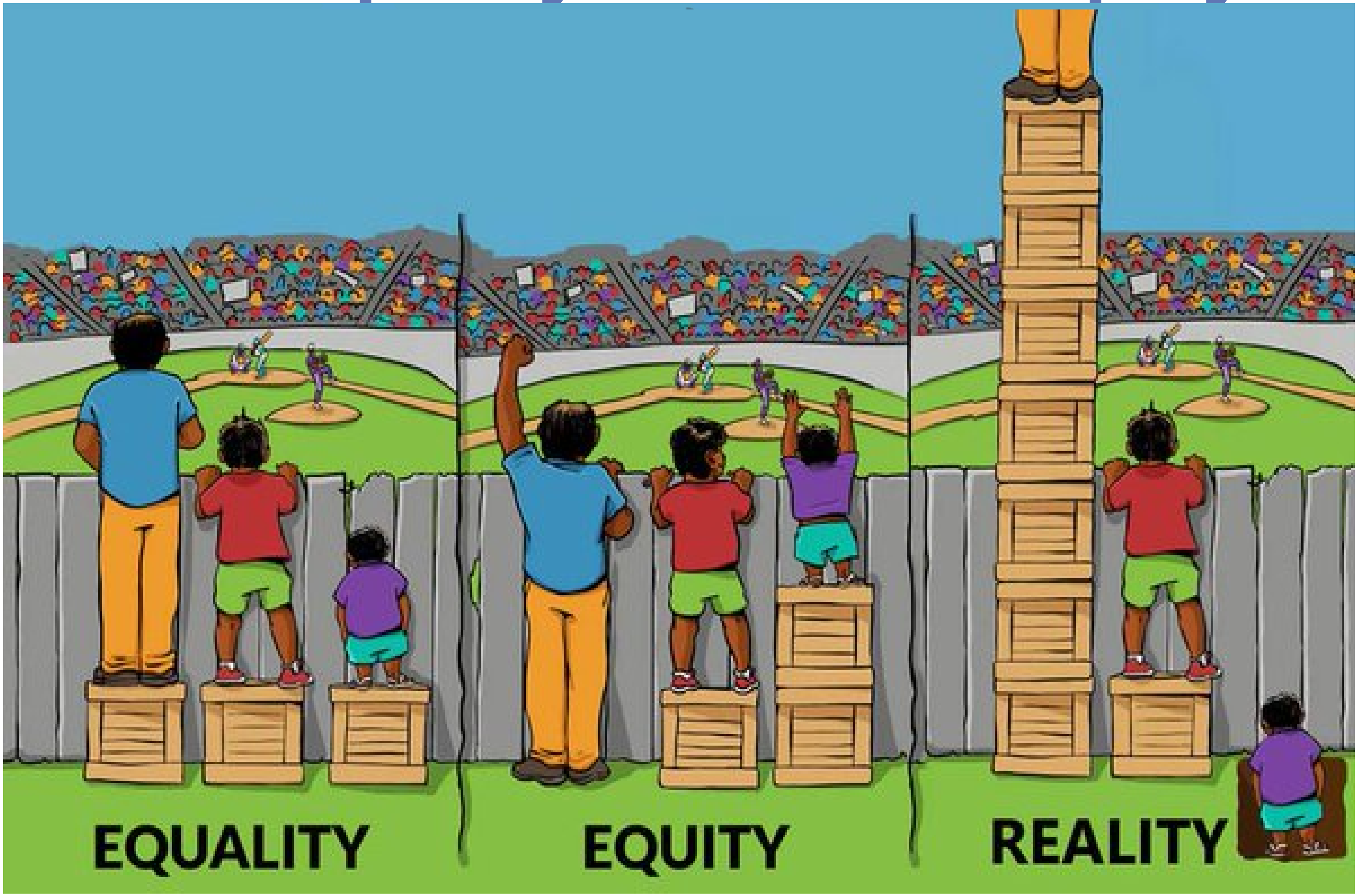
Epi222: Lecture 8
Analytic Issues in Health Disparities Research

Mahasin S. Mujahid, PhD, MS, FAHA
Associate Professor, Division of Epidemiology
Director, Epidemiology & Biostatistics MPH Program
University of California, Berkeley
February 28, 2017

Health Equality vs. Health Equity



Outline

- Background

- Two Analytic Challenges
  1. Confounding, Effect Modification, and Mediation
  2. Multi-level Models and Neighborhood Health Effects

ANALYTIC CHALLENGE #1

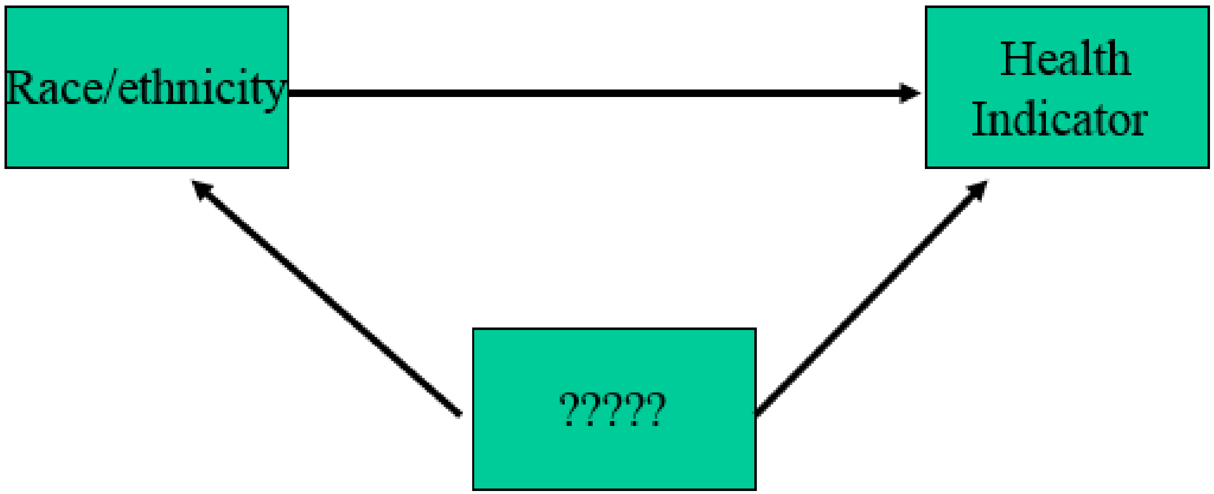
Confounding, Effect Modification, Mediation

Homework Question #1

- What are the different ways to account for SES in an analytic model when investigating racial/ethnic health disparities? (Hint: you should have three options). Discuss the interpretations/implications of each approach as it relates to the interest in understand health disparities by race/ethnicity.

Do Disparities Exist?

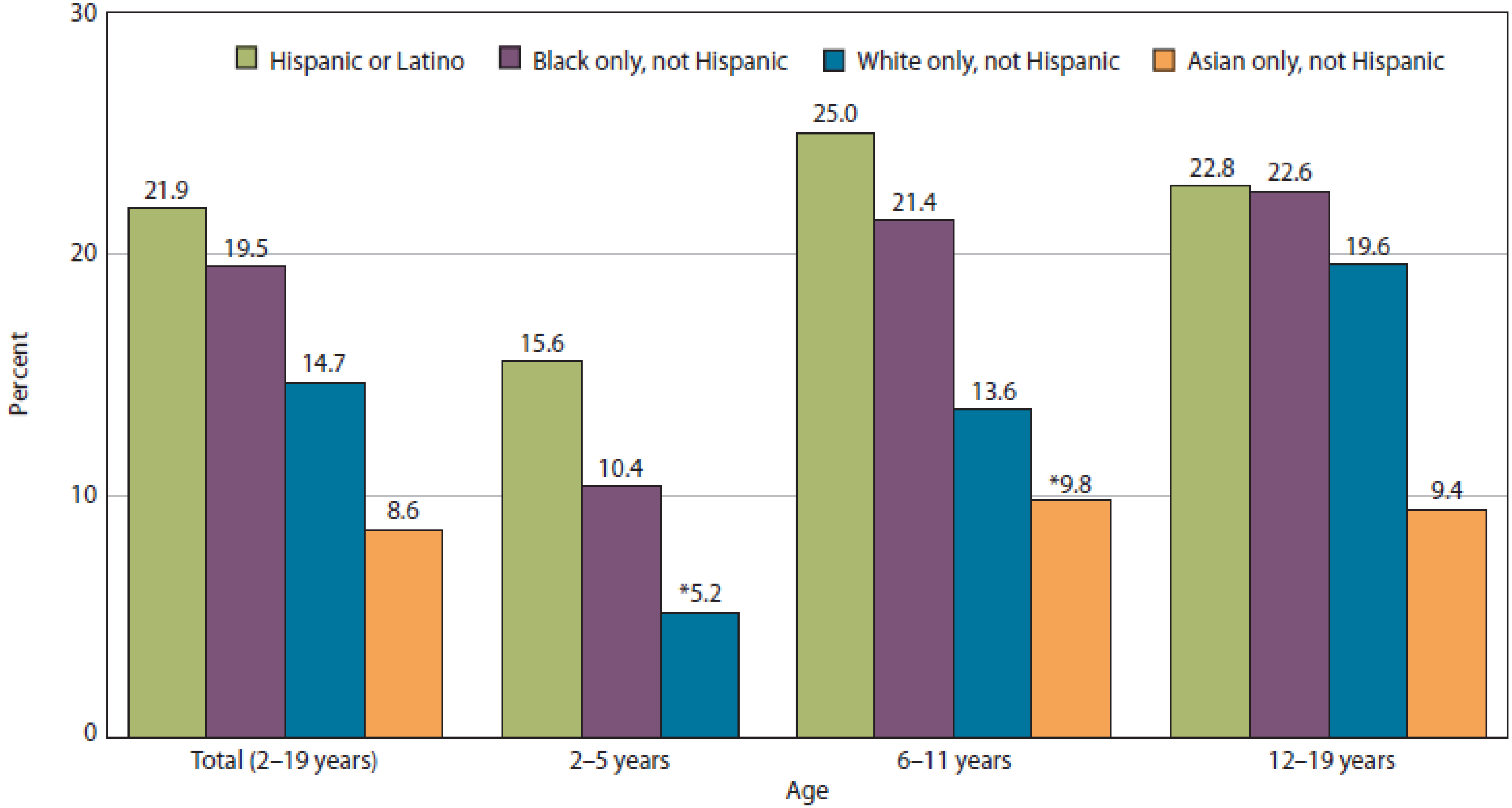
- Research Question: Are there differences in a particular health outcome by race/ethnicity



What are the basic set of confounders we should adjust for in an analytic model??

Example of Adjusting for Confounders

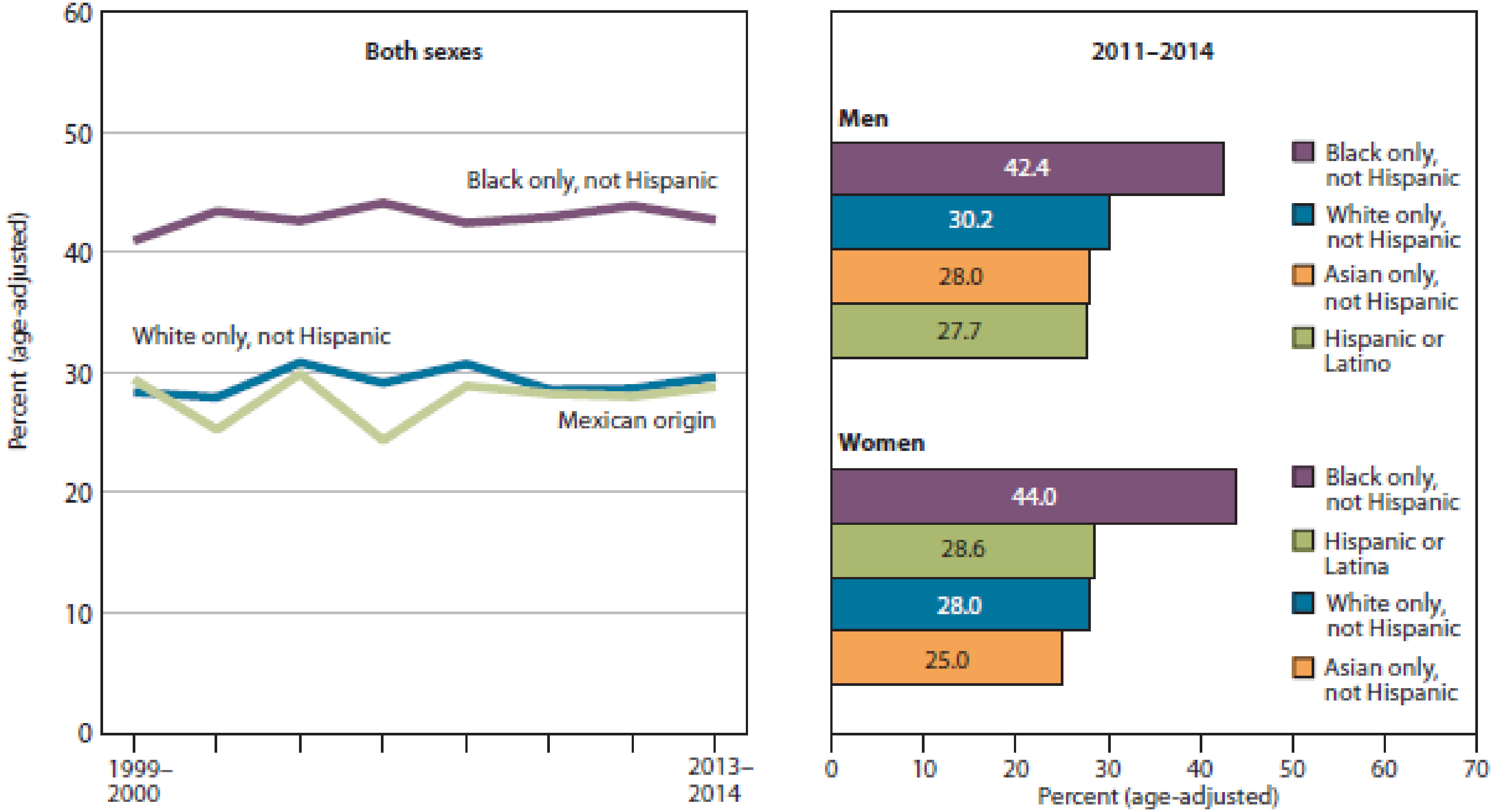
Figure 22. Obesity among children and adolescents aged 2–19 years, by age and race and Hispanic origin: United States, 2011–2014



CDC Chartbook 2015

Example of Adjusting for Confounders

Figure 23: Hypertension among adults aged 20 and over, by sex and race and Hispanic origin: United States, 1999–2000 through 2013–2014

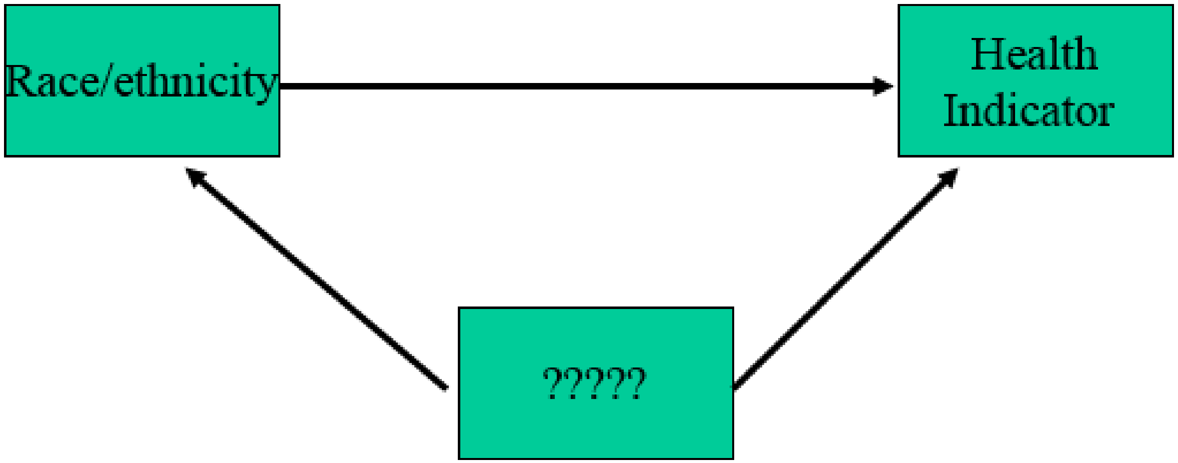


CDC Chartbook 2015

Do Disparities Exist?

- Research Question: Are there differences in a particular health outcome by race/ethnicity

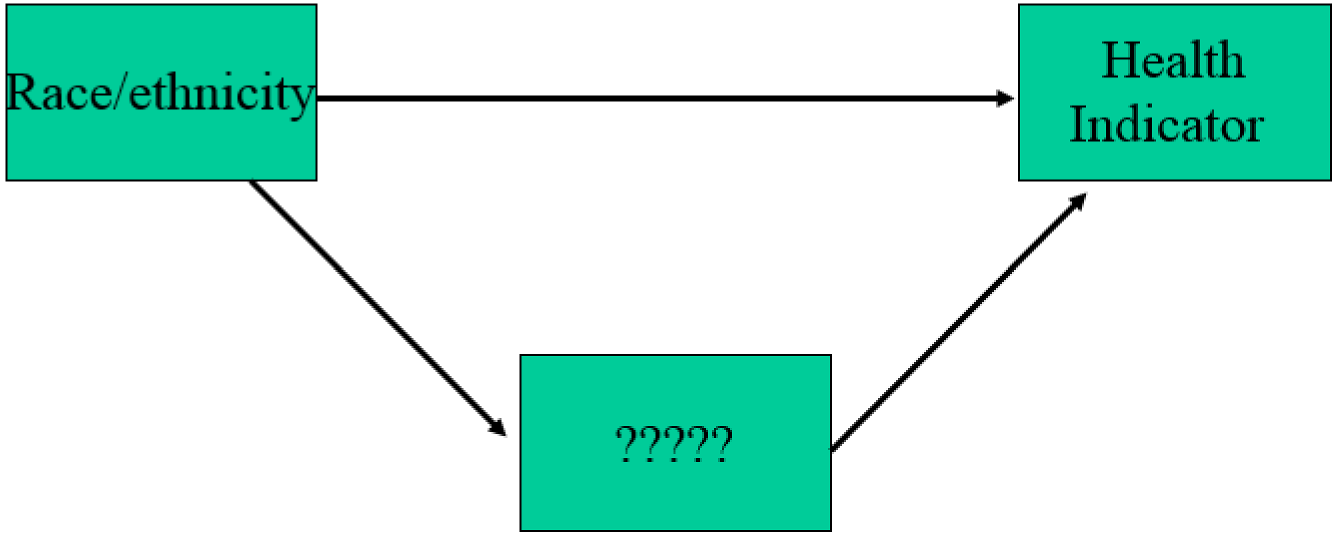
$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \varepsilon_i x_i$$



What are the basic set of confounders we should adjust for in an analytic model??

Why Do Racial/Ethnic Disparities Exist?

- Research Question: What are the mechanisms linking race/ethnicity and health



What are the mediators we should adjust for in an analytic model??

Approaches to Health Disparities Research

1. Genetic Model
2. Fundamental Cause Model
3. Pathways Model
4. Interaction Model

Diez Roux AV, 2012. Annual Review of Public Health

Explaining Black-White Differences in CHD: The REGARDS Study

Table 3. Total, Fatal, and Nonfatal Acute Coronary Heart Disease (CHD) Events

	Black vs White, Hazard Ratio (95% CI)	
	Men	**Women**
Total CHD		
Model 1[a]	1.15 (0.94-1.41)	1.48 (1.15-1.90)
Model 2[b]	1.04 (0.84-1.29)	1.25 (0.96-1.62)
Model 3[c]	0.93 (0.74-1.16)	0.95 (0.72-1.25)
Model 4[d]	0.87 (0.69-1.08)	0.90 (0.68-1.20)
Fatal CHD		
Model 1[a]	2.18 (1.55-3.06)	1.93 (1.23-3.03)
Model 2[b]	1.78 (1.24-2.56)	1.63 (1.02-2.62)
Model 3[c]	1.50 (1.02-2.19)	1.20 (0.73-1.97)
Model 4[d]	1.34 (0.91-1.96)	1.14 (0.69-1.99)
Nonfatal CHD		
Model 1[a]	0.81 (0.63-1.06)	1.31 (0.97-1.77)
Model 2[b]	0.78 (0.60-1.03)	1.10 (0.80-1.52)
Model 3[c]	0.72 (0.54-0.95)	0.85 (0.61-1.19)
Model 4[d]	0.68 (0.51-0.91)	0.81 (0.58-1.15)

[a] Adjusted for age and region of residence.
[b] Adjusted for model 1 covariates plus education level and income.
[c] Adjusted for model 2 covariates plus total cholesterol, high-density lipoprotein cholesterol, smoking status, systolic blood pressure, diabetes, and use of antihypertensive and statin medications.
[d] Adjusted for model 3 covariates plus body mass index, estimated glomerular filtration rate of less than 60 mL/min, log-transformed high-sensitivity C-reactive protein, and log-transformed albumin-to-creatinine ratio.

Safford MM et al., 2012: Journal of the American Medical Association

Factors that Mediate Racial/Ethnic Disparities in US Fetal Death Rates

- Objective: To determine the importance of socioeconomic factors, maternal comorbid conditions, antepartum and intrapartum complication of pregnancy, and fetal factors in mediating racial disparities in fetal deaths

Lorch SA, et al., 2012 AJPH

Factors that Mediate Racial/Ethnic Disparities in US Fetal Death Rates

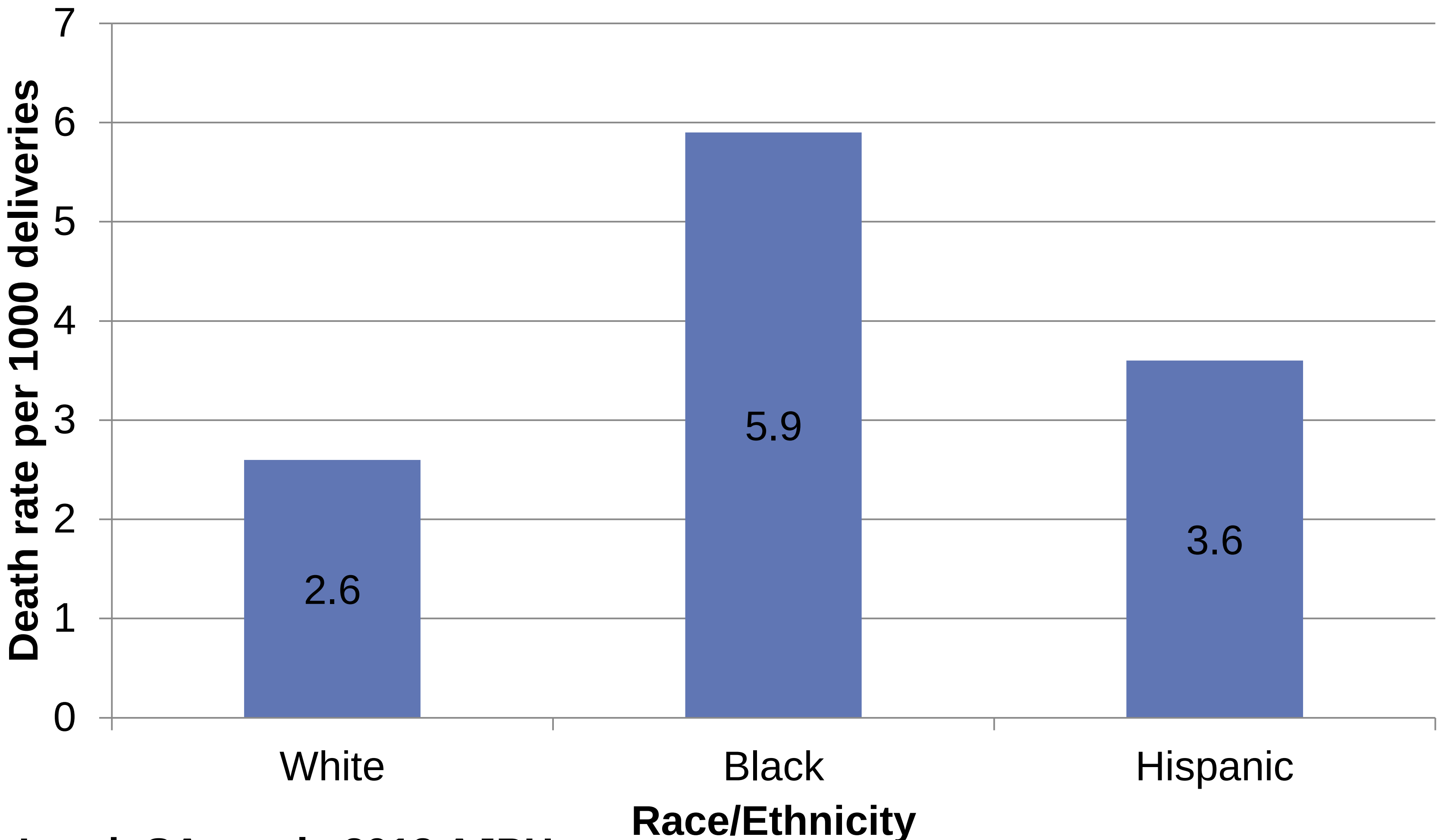
- Study Design: Retrospective Cohort Study of hospital-based deliveries (gestational age: 22-44 weeks) in California, Missouri, and Pennsylvania

- Study Period: 1994-2005

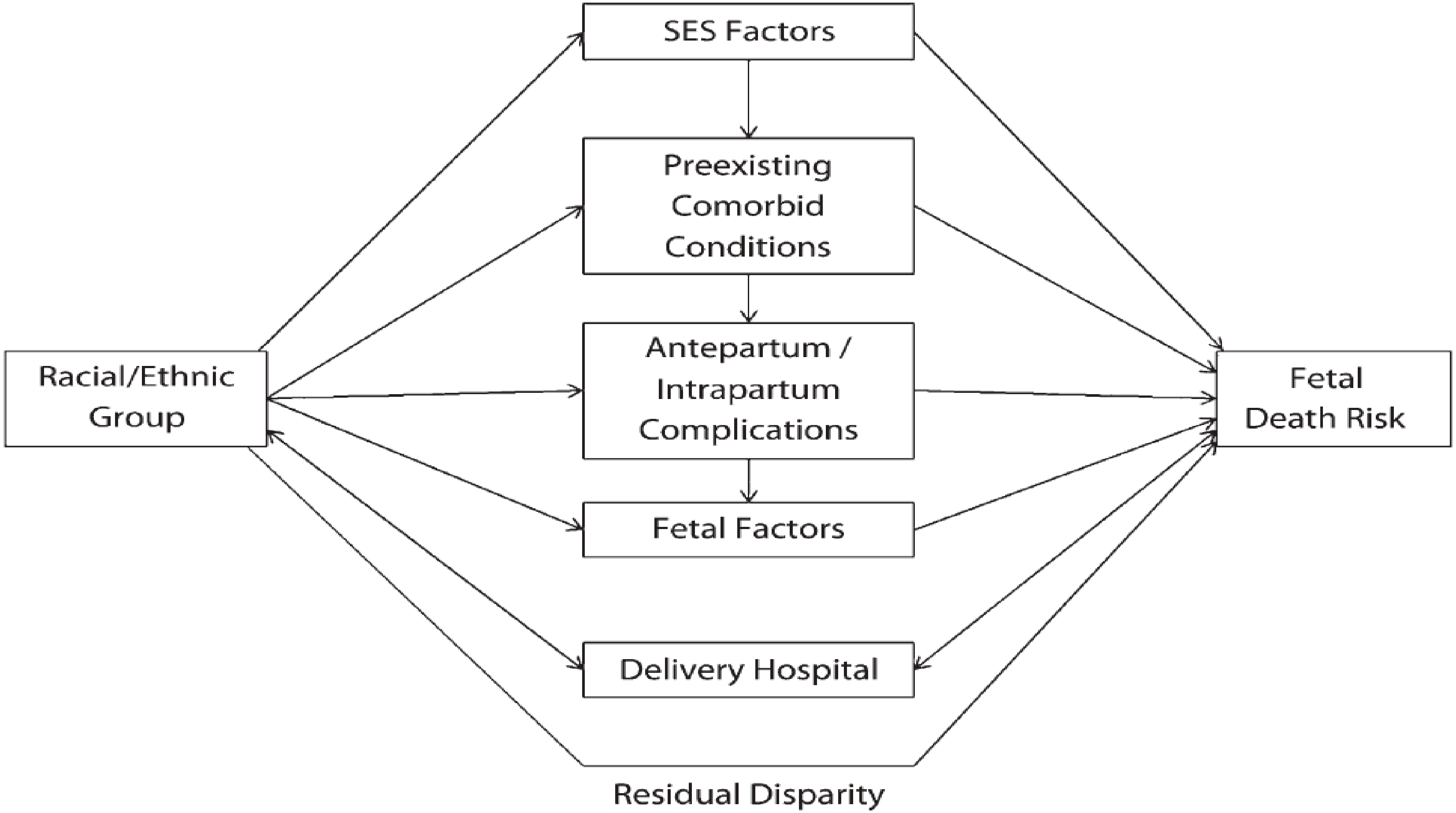
- Sample size: N=7,104,674 births

Lorch SA, et al., 2012 AJPH

Fetal Death Rate by Race/Ethnicity



Lorch SA, et al., 2012 AJPH



Note. SES = socioeconomic status. The residual disparity in fetal deaths, or fetal deaths unexplained by any of the included factors, is shown as the bottom pathway.

FIGURE 1—Hierarchical conceptual framework: racial/ethnic differences in fetal death, California, Missouri, Pennsylvania, 1993–2005.

Lorch SA, et al., 2012 AJPH

Baron and Kenny Mediation Framework

- Mediation testing. On the basis of Baron and Kenny's framework, a set of factors may mediate the relationship between racial/ethnic group and fetal death risk if the following conditions are met:
  1. Racial/ethnic group is associated with fetal death risk.
  2. Racial/ethnic group is associated with a set of potential mediating factors.
  3. A set of potential mediating factors are associated with the risk of fetal death.
  4. Including both racial/ethnic group and the set of mediating factors in a model changes the association between fetal death and racial/ethnic group observed in condition 1.

Lorch SA, et al., 2012 AJPH

CONDITION 1

- In unadjusted analysis, all 3 racial/ethnic groups had higher odds of a fetal death than did non-Hispanic White women
  - Blacks: OR = 2.24; 95% CI = 2.08, 2.42;
  - Hispanics: OR = 1.37; 95% CI = 1.28, 1.47;
  - Asians: OR = 1.18; 95% CI = 1.09, 1.28

Lorch SA, et al., 2012 AJPH

CONDITION 2

TABLE 2—AORs by Selected Characteristics With Distribution of Characteristics: Racial/Ethnic Differences in Fetal Death, California, Missouri, Pennsylvania, 1993–2005

Characteristic	Fetal Death, AOR (95% CI)	White (n = 3 267 285), %	Black (n = 541 400), %	Hispanic (n = 2 461 374), %	Asian (n = 531 164), %
Socioeconomic factors					
Maternal age, y					
< 18	0.97 (0.90, 1.03)	2.43	7.31	5.59	1.45
18–35	1.00 (Ref)	83.54	84.13	86.28	81.42
> 35	1.61 (1.55, 1.67)	14.03	8.56	8.12	17.13
Insurance status					
Fee for service	1.00 (Ref)	16.00	5.53	2.65	4.83
Health maintenance organization	1.08 (0.98, 1.19)	53.63	31.93	29.66	65.91
Public	1.31 (1.19, 1.45)	23.95	54.57	63.52	24.34
Other	0.95 (0.80, 1.13)	4.13	5.71	0.63	1.11
Uninsured	2.02 (1.78, 2.30)	2.11	2.05	3.50	3.78
Missing	1.31 (0.90, 1.89)	0.19	0.21	0.04	0.04

Lorch SA, et al., 2012 AJPH

CONDITION 3

TABLE 2—AORs by Selected Characteristics With Distribution of Characteristics: Racial/Ethnic Differences in Fetal Death, California, Missouri, Pennsylvania, 1993–2005

Characteristic	Fetal Death, AOR (95% CI)	White (n = 3 267 285), %	Black (n = 541 400), %	Hispanic (n = 2 461 374), %	Asian (n = 531 164), %
Socioeconomic factors					
Maternal age, y					
< 18	0.97 (0.90, 1.03)	2.43	7.31	5.59	1.45
18-35	1.00 (Ref)	83.54	84.13	86.28	81.42
> 35	1.61 (1.55, 1.67)	14.03	8.56	8.12	17.13
Insurance status					
Fee for service	1.00 (Ref)	16.00	5.53	2.65	4.83
Health maintenance organization	1.08 (0.98, 1.19)	53.63	31.93	29.66	65.91
Public	1.31 (1.19, 1.45)	23.95	54.57	63.52	24.34
Other	0.95 (0.80, 1.13)	4.13	5.71	0.63	1.11
Uninsured	2.02 (1.78, 2.30)	2.11	2.05	3.50	3.78
Missing	1.31 (0.90, 1.89)	0.19	0.21	0.04	0.04

Lorch SA, et al., 2012 AJPH

TABLE 2—AORs by Selected Characteristics With Distribution of Characteristics: Racial/Ethnic Differences in Fetal Death, California, Missouri, Pennsylvania, 1993–2005

Characteristic	Fetal Death, AOR (95% CI)	White (n = 3 267 285), %	Black (n = 541 400), %	Hispanic (n = 2 461 374), %	Asian (n = 531 164), %
Trimester of first prenatal care visit, mo					
0-3	1.00 (Ref)	87.61	75.57	79.74	85.64
4-6	1.18 (1.11, 1.25)	8.57	15.97	15.26	10.76
7-9	1.15 (1.02, 1.28)	1.54	3.11	3.17	2.00
Missing	1.98 (1.60, 2.45)	2.28	5.35	1.83	1.60
Education					
< high school	1.06 (1.01, 1.11)	11.06	21.64	51.68	11.85
≥ high school	1.00 (Ref)	87.68	75.71	46.56	86.20
Missing	3.83 (2.99, 4.89)	1.26	2.66	1.76	1.95
Preexisting comorbid conditions					
Gestational diabetes	0.88 (0.82, 0.94)	3.42	3.14	4.49	6.78
Diabetes mellitus	3.17 (2.85, 3.53)	0.55	0.90	0.81	0.62
Chronic hypertension	2.20 (1.99, 2.43)	0.72	1.51	0.37	0.53
Renal disease	2.51 (1.92, 3.28)	0.12	0.14	0.08	0.08

Lorch SA, et al., 2012 AJPH

TABLE 2—AORs by Selected Characteristics With Distribution of Characteristics: Racial/Ethnic Differences in Fetal Death, California, Missouri, Pennsylvania, 1993–2005

Characteristic	Fetal Death, AOR (95% CI)	White (n = 3 267 285), %	Black (n = 541 400), %	Hispanic (n = 2 461 374), %	Asian (n = 531 164), %
Fetal characteristics					
Birth weight categories, g					
400-999	1963.00 (1710.00, 2254.00)	0.47	1.32	0.48	0.40
1000-1499	579.00 (505.00, 663.00)	0.63	1.37	0.56	0.54
1500-1999	195.40 (171.10, 223.10)	1.34	2.65	1.14	1.25
2000-2499	27.20 (24.10, 30.70)	4.06	7.55	3.65	4.62
2500-2999	3.68 (3.30, 4.10)	14.27	22.57	15.52	21.61
3000-3499	1.17 (1.05, 1.30)	35.89	37.33	38.88	42.60
3500-3999	0.81 (0.72, 0.90)	31.51	21.22	29.83	23.22
4000-4499	1.00 (Ref)	10.01	5.06	8.41	4.99
≥ 4500	5.04 (4.36, 5.84)	1.80	0.87	1.52	0.77
Missing	1129.00 (519.00, 2451.00)	0.02	0.06	0.00	0.01
Gestational status					
Multiple gestation	0.59 (0.54, 0.64)	3.60	3.71	2.02	2.31
Small-for-gestational-age	4.85 (4.65, 5.06)	7.90	14.75	8.42	11.86

Lorch SA, et al., 2012 AJPH

Modeling Approach

- Sequential logistic regression models
  - Model A: Unadjusted
  - Model B: Socioeconomic factors
  - Model C: Preexisting maternal conditions
  - Model D: Antepartum and intrapartum complications
  - Model E: Fetal factors
  - Model F: Delivery hospital
- $Mediation\ \% = \dfrac{coef.for\ racial,ethnic\ groups(prior) - coef.for\ group\ (successive)}{coef.for\ racial,ethnic\ group(crude)}$

Non-parametric bootstrapping to determine confidence intervals for mediation %

Lorch SA, et al., 2012 AJPH

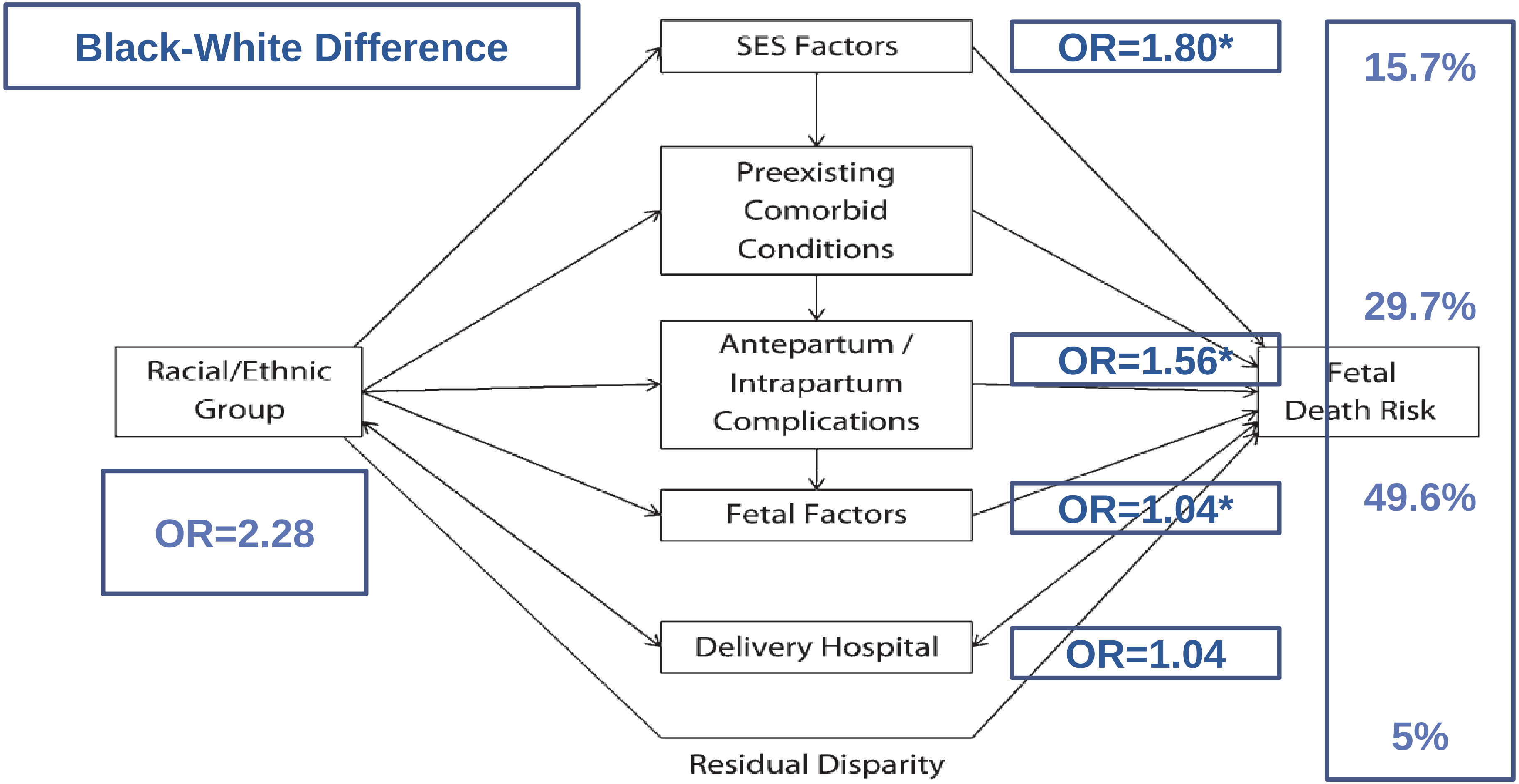
CONDITION 4

TABLE 3—The Impact of Potential Mediating Factors: Racial/Ethnic Differences in Fetal Death, California, Missouri, Pennsylvania, 1993–2005

Mediating Factor	Black, % (95% CI)	Hispanic, % (95% CI)	Asian, % (95% CI)
Socioeconomic factors	15.7 (10.6, 20.1)	35.8 (25.8, 46.2)	37.8 (13.3, 65.4)
Antepartum and intrapartum complications of pregnancy	29.7 (24.7, 34.7)	. . .	62.2 (62.2, 10.0)
Fetal factors	49.6 (42.7, 54.7)	. . .	. . .
Residual disparity	5.0	64.2	3.4

Note. CI = confidence interval. Maternal preexisting comorbid conditions did not mediate a statistically significant proportion of fetal death disparity for any racial/ethnic group. Ellipses indicate that the percentage of fetal death disparity mediated by set of factors is not statistically significant for specific racial/ethnic group.

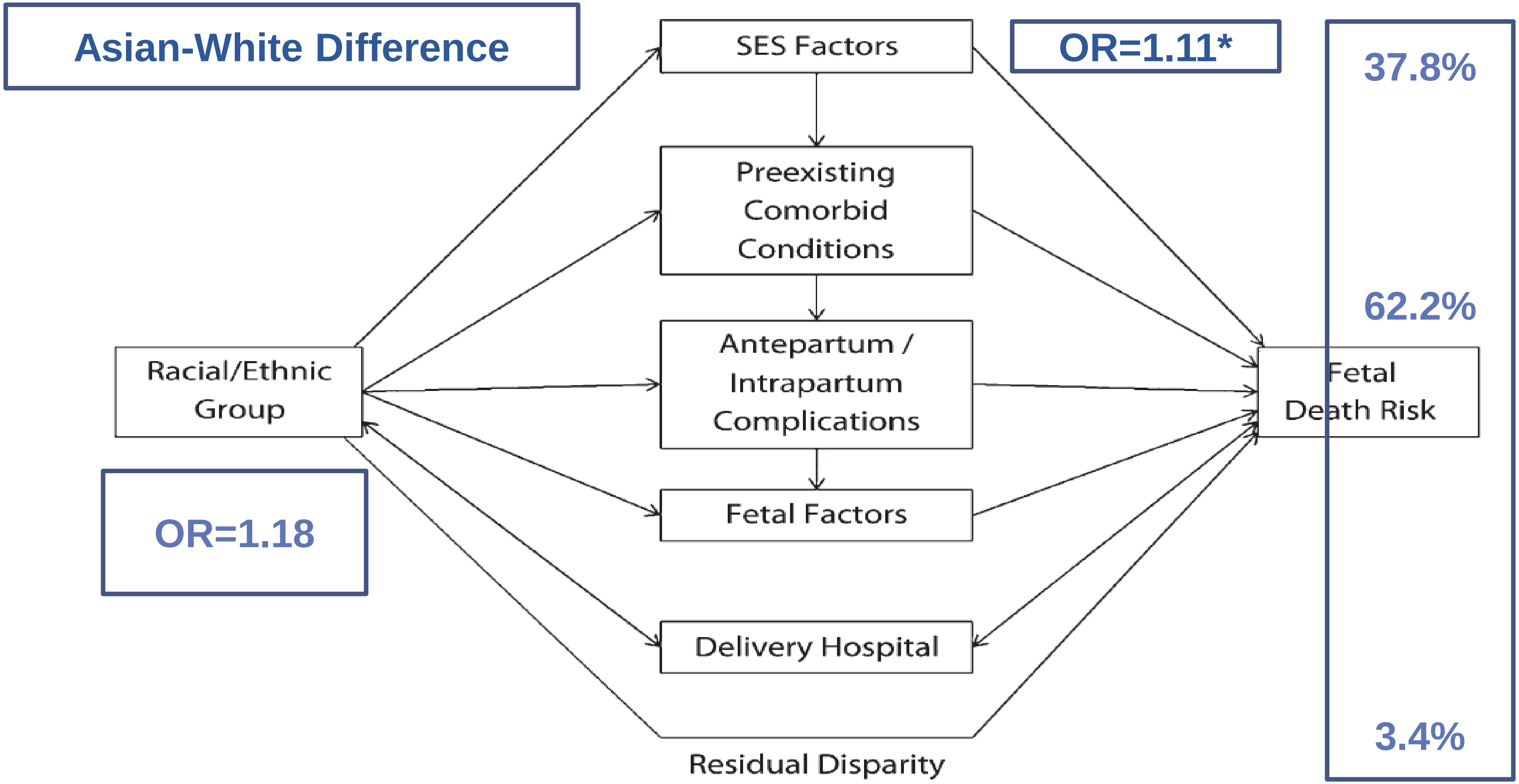
Lorch SA, et al., 2012 AJPH



Note. SES = socioeconomic status. The residual disparity in fetal deaths, or fetal deaths unexplained by any of the included factors, is shown as the bottom pathway.

FIGURE 1—Hierarchical conceptual framework: racial/ethnic differences in fetal death, California, Missouri, Pennsylvania, 1993–2005.

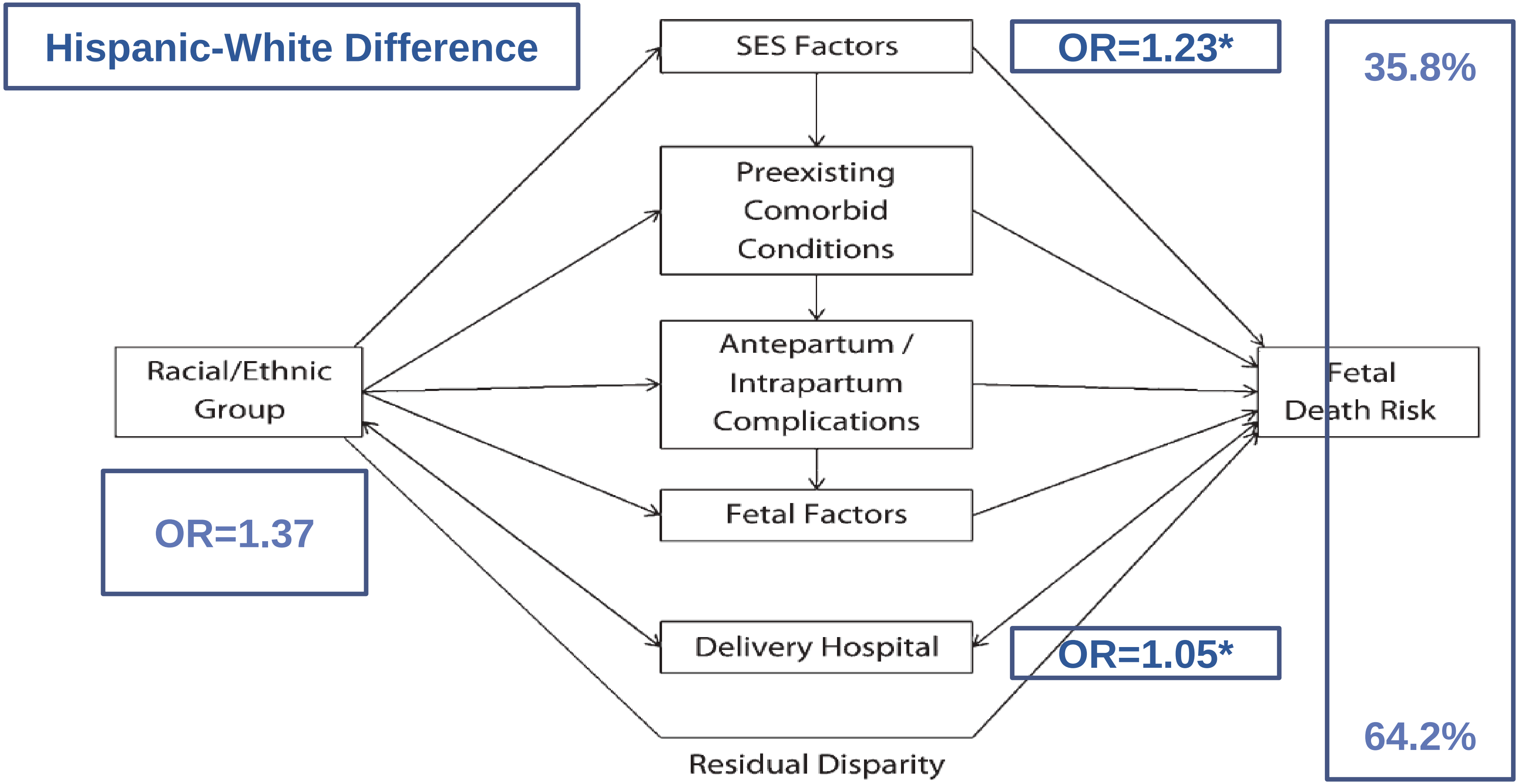
Lorch SA, et al., 2012 AJPH



Note. SES = socioeconomic status. The residual disparity in fetal deaths, or fetal deaths unexplained by any of the included factors, is shown as the bottom pathway.

FIGURE 1—Hierarchical conceptual framework: racial/ethnic differences in fetal death, California, Missouri, Pennsylvania, 1993–2005.

Lorch SA, et al., 2012 AJPH

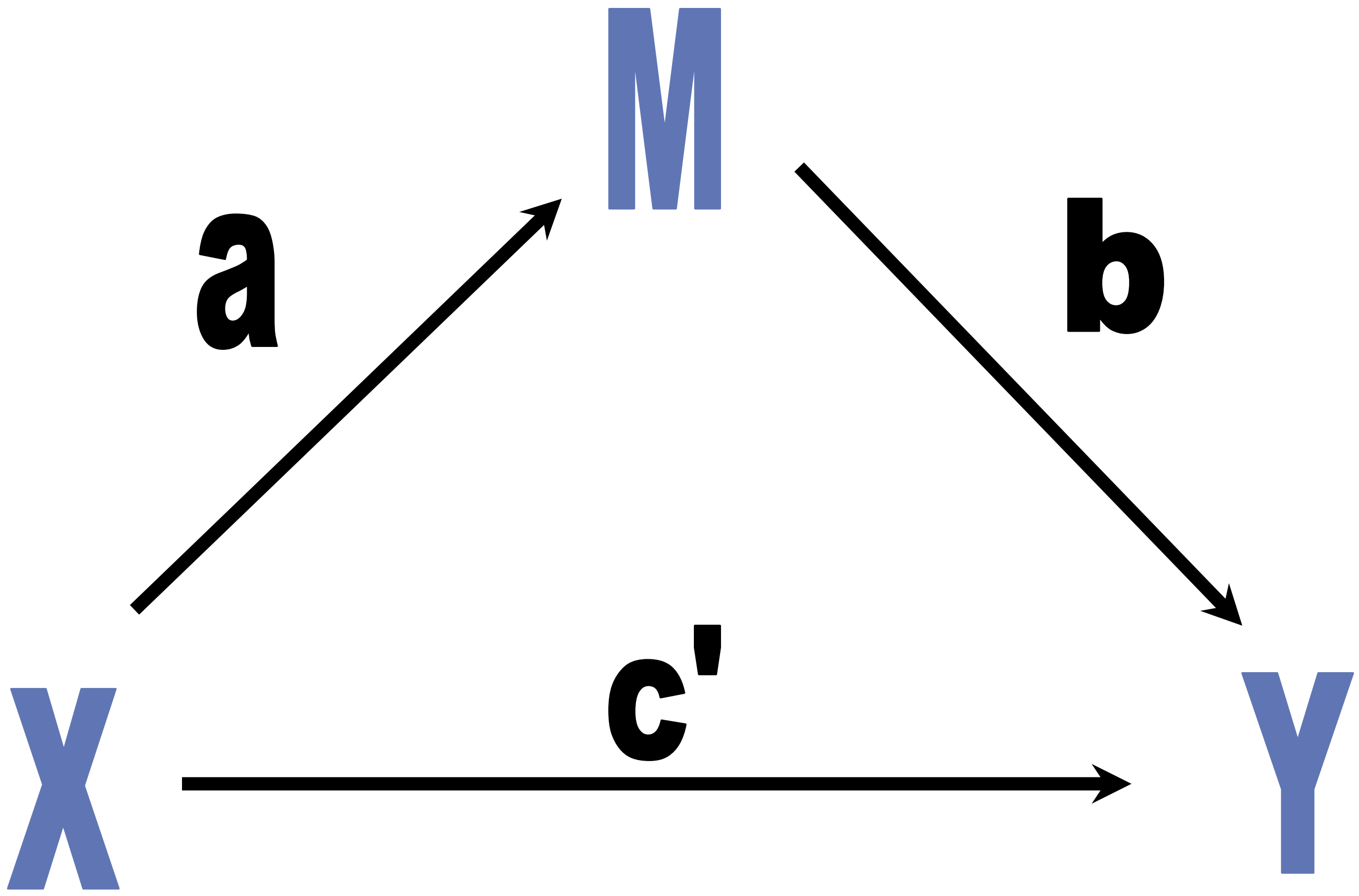


Note. SES = socioeconomic status. The residual disparity in fetal deaths, or fetal deaths unexplained by any of the included factors, is shown as the bottom pathway.

FIGURE 1—Hierarchical conceptual framework: racial/ethnic differences in fetal death, California, Missouri, Pennsylvania, 1993–2005.

Lorch SA, et al., 2012 AJPH

The Mediation Model



The Four Step Approach

- Step 1: X → Y (test path *c*)
- Step 2: X → M (test path *a*)
- Step 3: M (and X) → Y (test path *b*)
- Step 4: X (and M) → Y (test path *c'*)

Note that Steps 3 and 4 use the same regression equation

Problems with Basic Mediation Analyses

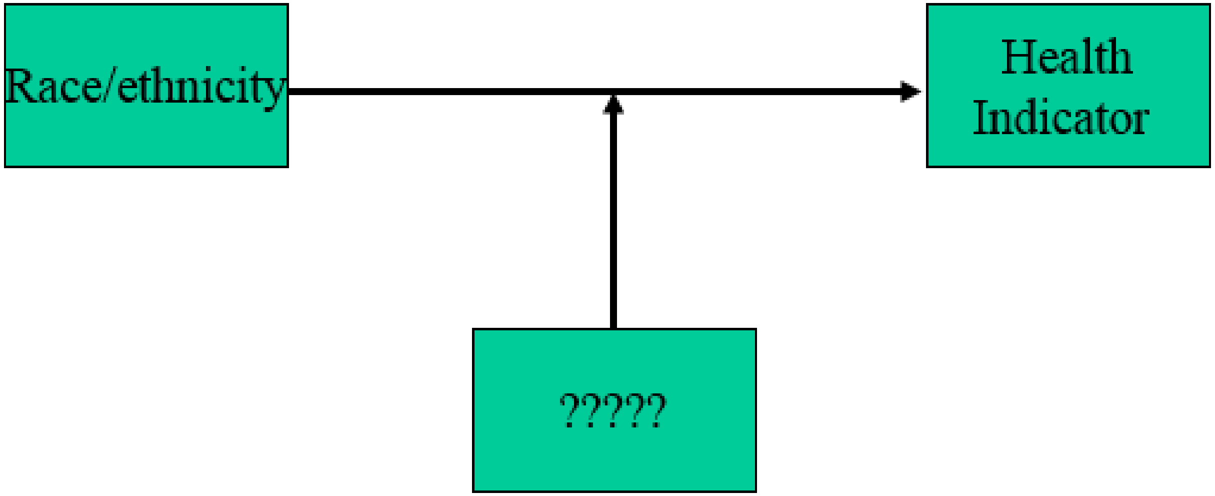
- Linearity
- Normal Distribution of Errors
  - Interval level of measurement of M and Y
- Equal Error Variance
- Independence
  - No clustering
- **X and M Do Not Interact to Cause Y**

Causal Assumptions

- No Measurement Error
  - for M and X
- No Omitted Variables
  - all common causes of M and Y, X and M, and X and Y measured and controlled
- No Reverse Causal Effects
  - M and Y not cause X
  - Y may not cause M

How do we eliminate racial/ethnic health disparities?

- Research Question: Does one size fit all? Does the magnitude of racial/ethnic differences in health depend on a third variable



What are the effect modifiers we should test for???

Problems with Effect Modification

- Effect modification is not the same as statistical interaction
  - Should you test for all two-way interactions in a model?
- Effect modification is not a bias or threat to internal validity
  - There are power issues in testing for interactions
  - Additive vs. multiplicative interactions
- Should we really expect any treatment to have the same effect for all?

Example of Effect Modification

- Objective: To determine whether racial/ethnic differences in adequate and excessive gestational weight gain vary by pre-pregnancy weight class
  - Three study outcomes: inadequate weight gain, excessive weight gain

Headen IH, Mujahid MS et al., 2015 Matern Child Health J

Example of Effect Modification

- Study Design: Prospective Cohort Study, US Bureau of Labor Statistics National Longitudinal Survey of Youth (NLSY)

- Study Period: 1979-2010 (annual follow-up through 1994 and biannual thereafter)

- Sample size: N=6,849 births

Headen IH, Mujahid MS et al., 2015 Matern Child Health J

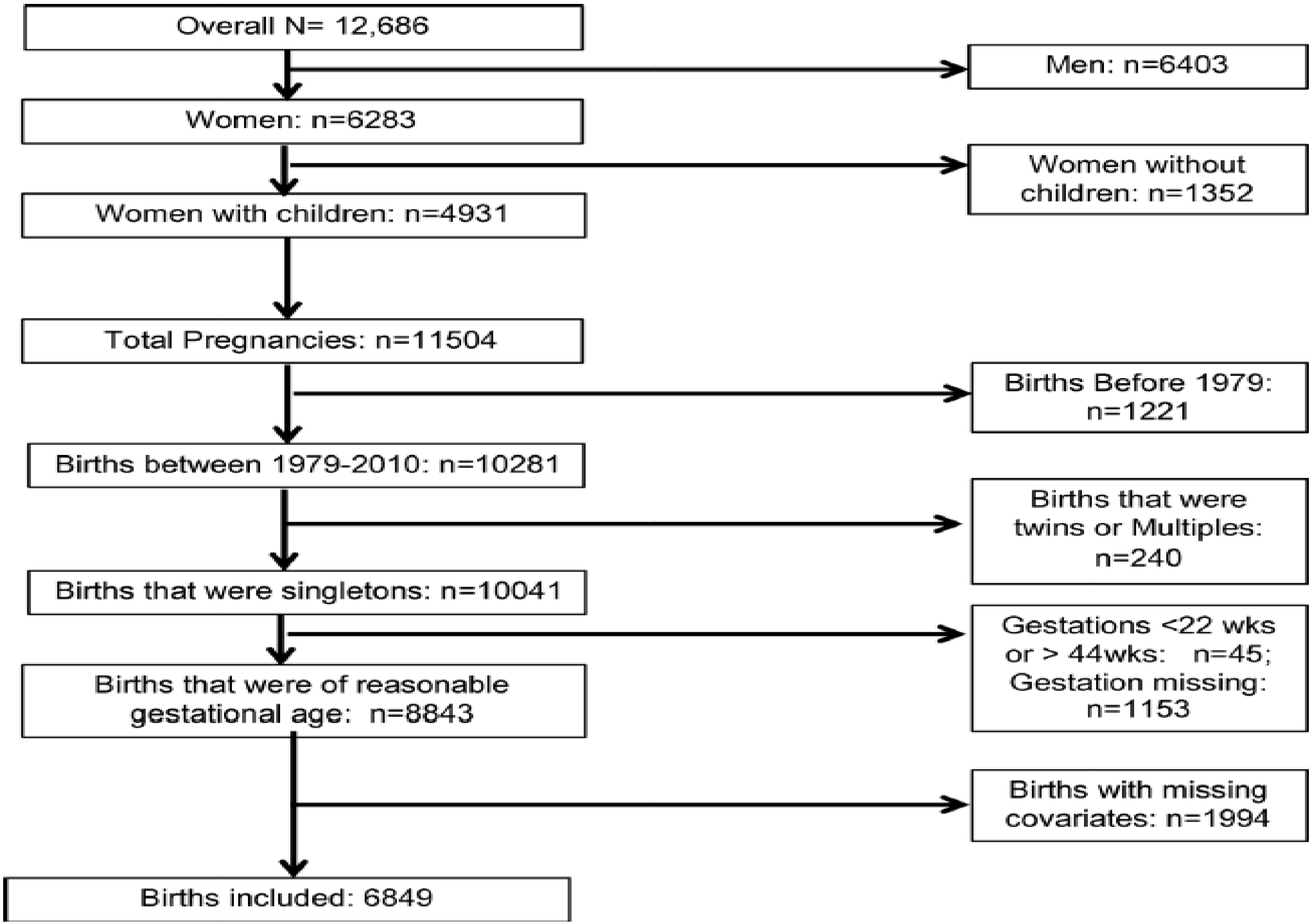
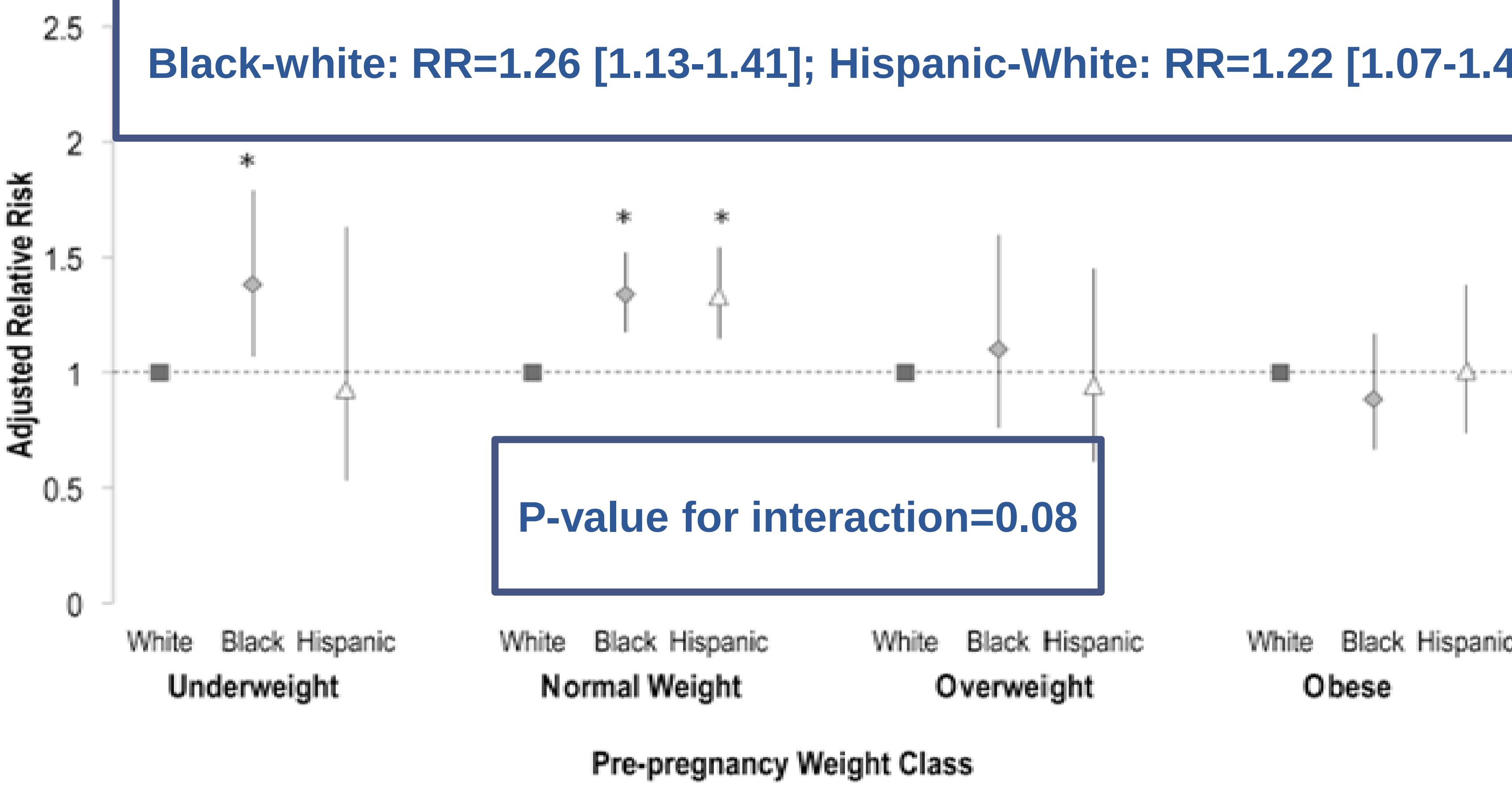


Figure 1. Observations from the NLSY79 Complete Cohort Remaining in Analytic Sample Based on Inclusion and Exclusion Criteria

Race/Ethnicity, Pre-pregnancy BMI and Inadequate Gain

Black-white: RR=1.26 [1.13-1.41]; Hispanic-White: RR=1.22 [1.07-1.40]



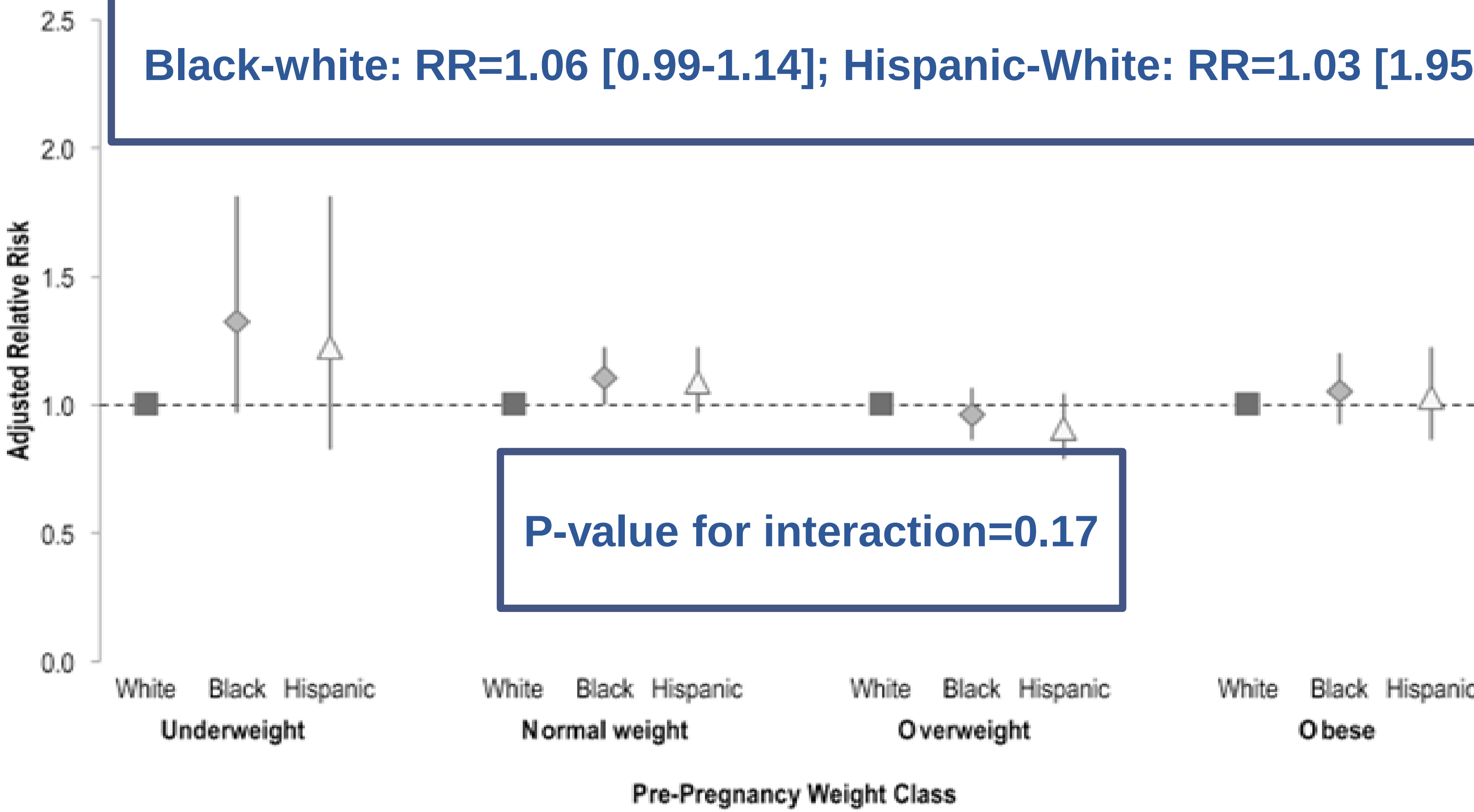
P-value for interaction=0.08

Figure 2. Racial Differences in Inadequate Gestational Weight Gain by Pre-pregnancy Weight Class (births n= 3788; moms n= 2440)[a]

Headen IH, Mujahid MS et al., 2015 Matern Child Health J

Race/Ethnicity, Pre-pregnancy BMI and Inadequate Gain

Black-white: RR=1.06 [0.99-1.14]; Hispanic-White: RR=1.03 [1.95-1.12]



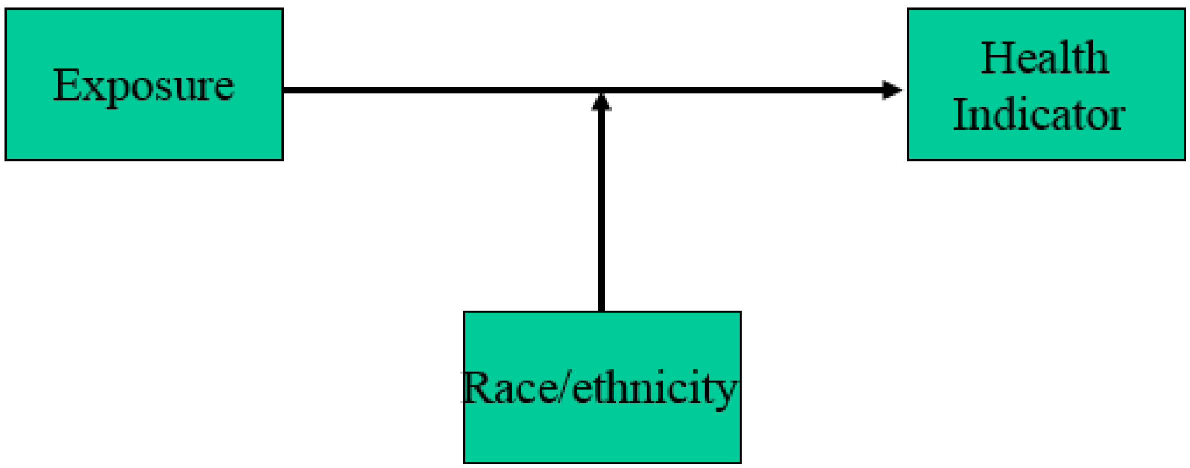
P-value for interaction=0.17

Figure 3. Racial Differences in Excessive Gestational Weight Gain by Pre-Pregnancy Weight Class (births n=5128; moms n=3190)[a]

Headen IH, Mujahid MS et al., 2015 Matern Child Health J

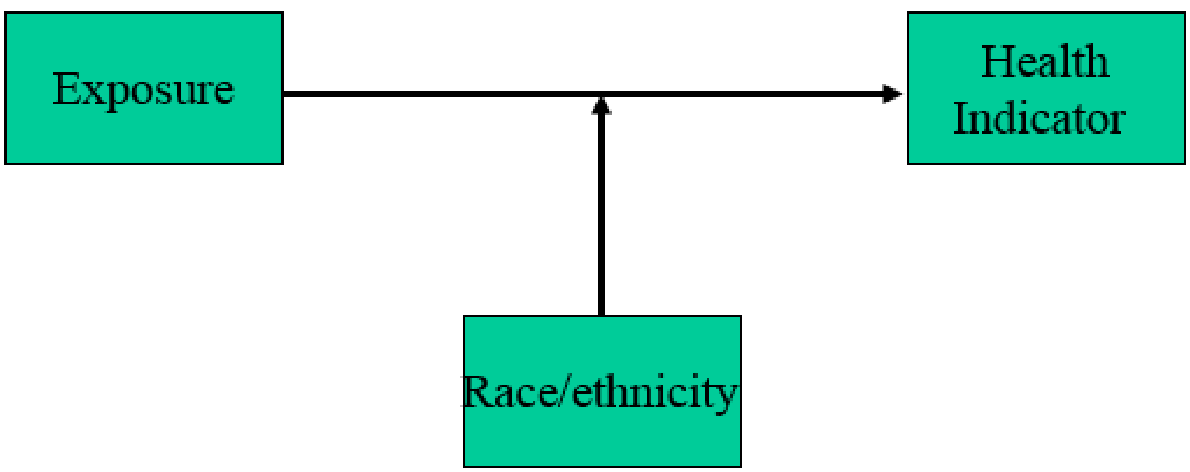
How do we eliminate racial/ethnic health disparities?

- Research Question: Does a exposure-health outcome association vary by racial/ethnic groups



How do we eliminate racial/ethnic health disparities?

- Why might we think the association between an exposure and an outcome be differential by race/ethnicity



Measurement error of exposures
Lack of equivalence of exposures values
Differential mechanisms

ANALYTIC CHALLENGE #2

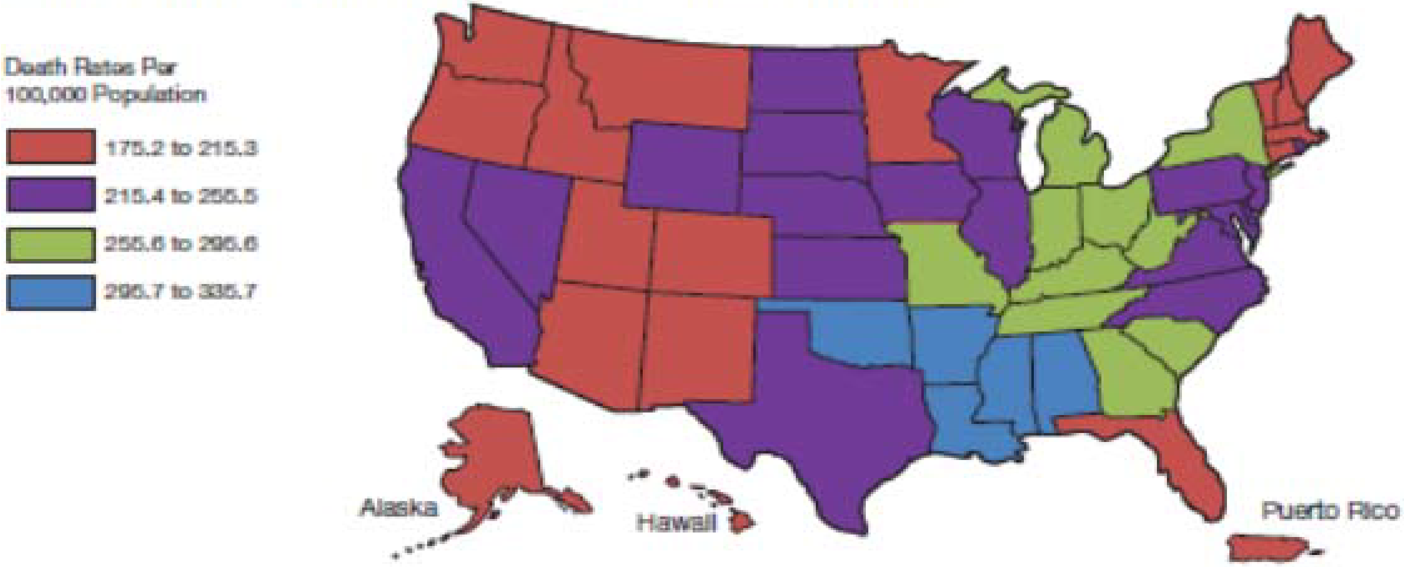
Multi-level Models and Neighborhood Health Effects

Homework Question #2

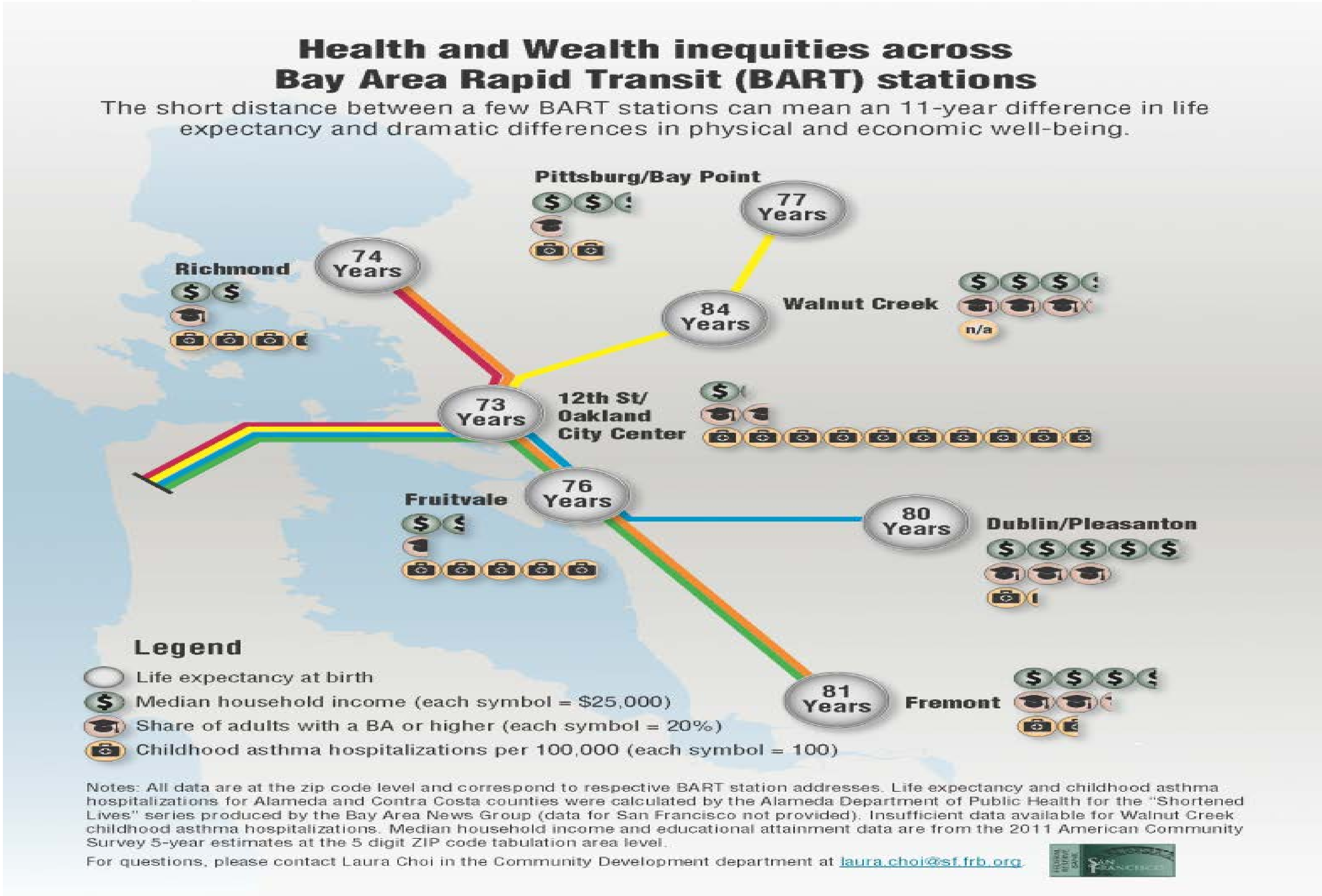
- **Select a research question investigating associations between multi-level social factors (operating at least two levels) and any health outcome. State the exposures and outcomes, the additional study covariates that would be included in the analytic model, and a discussion of the analytic considerations in an multi-level investigation**

CVD Mortality by Place

Major Cardiovascular Disease Age-Adjusted Death Rates by State



Go A S et al. Circulation. 2014;129:e28-e292



Health and Wealth inequities across
Bay Area Rapid Transit (BART) stations
The short distance between a few BART stations can mean an 11-year difference in life expectancy and dramatic differences in physical and economic well-being.
Pittsburg/Bay Point
77 Years
Richmond
74 Years
84 Years
Walnut Creek
n/a
73 Years
12th St/ Oakland City Center
Fruitvale
76 Years
80 Years
Dublin/Pleasanton
81 Years
Fremont
Legend
Life expectancy at birth
Median household income (each symbol = $25,000)
Share of adults with a BA or higher (each symbol = 20%)
Childhood asthma hospitalizations per 100,000 (each symbol = 100)
Notes: All data are at the zip code level and correspond to respective BART station addresses. Life expectancy and childhood asthma hospitalizations for Alameda and Contra Costa counties were calculated by the Alameda Department of Public Health for the "Shortened Lives" series produced by the Bay Area News Group (data for San Francisco not provided). Insufficient data available for Walnut Creek childhood asthma hospitalizations. Median household income and educational attainment data are from the 2011 American Community Survey 5-year estimates at the 5 digit ZIP code tabulation area level.
For questions, please contact Laura Choi in the Community Development department at laura.choi@sf.frb.org.

Rockridge

Sources: www.transcolation.org, www.csua.berekeley.edu.

West Oakland

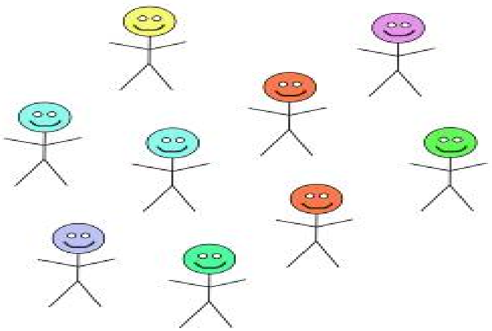
(www.eirikjohnson.com/westoakland

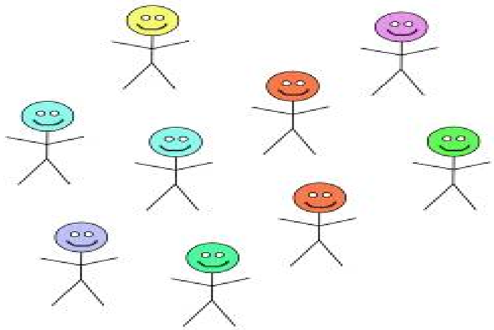
Counterfactual Question

- Suppose W. Oakland has double the mortality rate compared to Rockridge
- Does this mean <u>your</u> mortality risk would increase if you moved from Rockridge to W. Oakland?
- What if W. Oakland has twice the rate of poverty compared to Rockridge (and being poor doubles your mortality risk)
- But you yourself are not poor?

Adapted from I. Kawachi

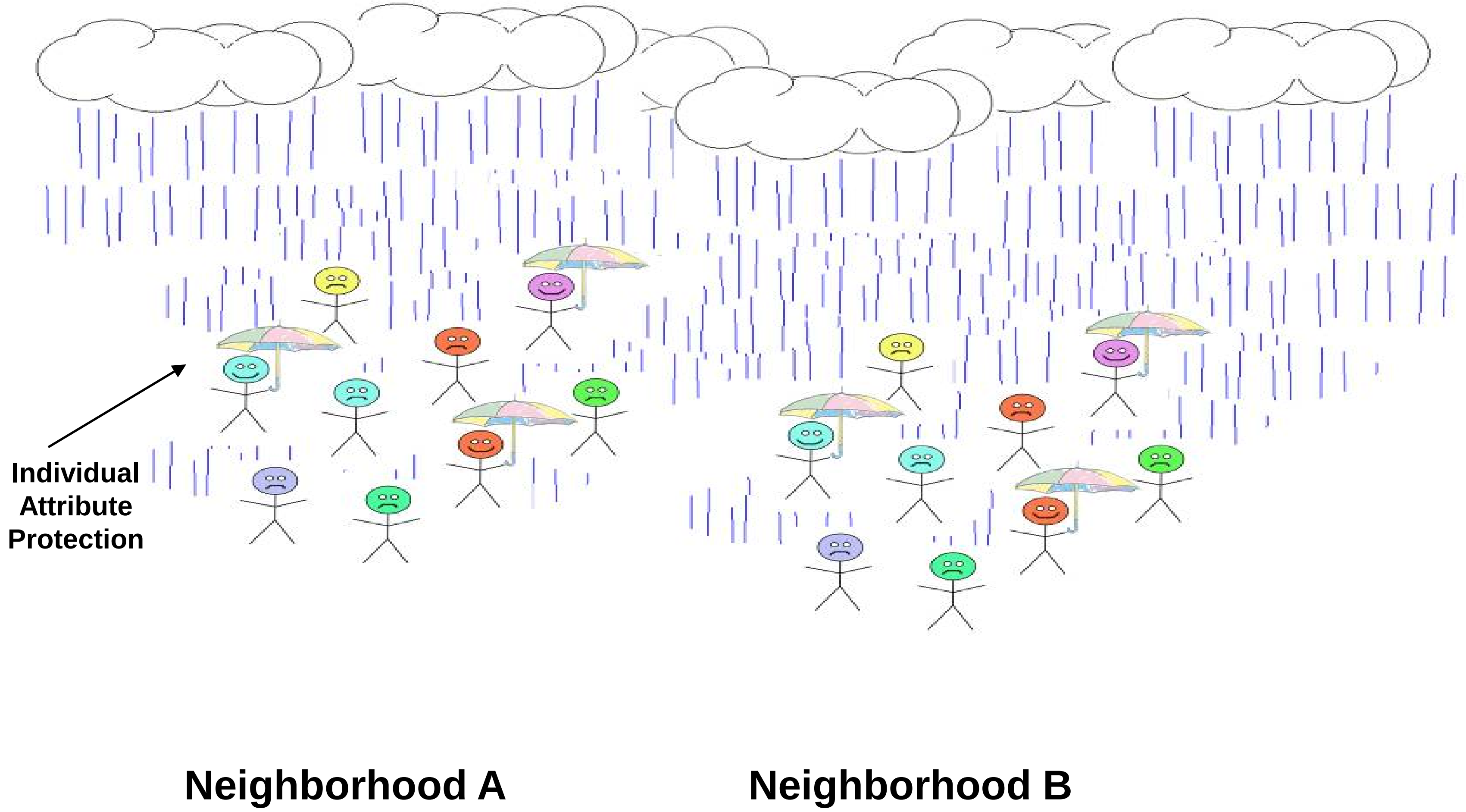
Individual and Neighborhood Effects



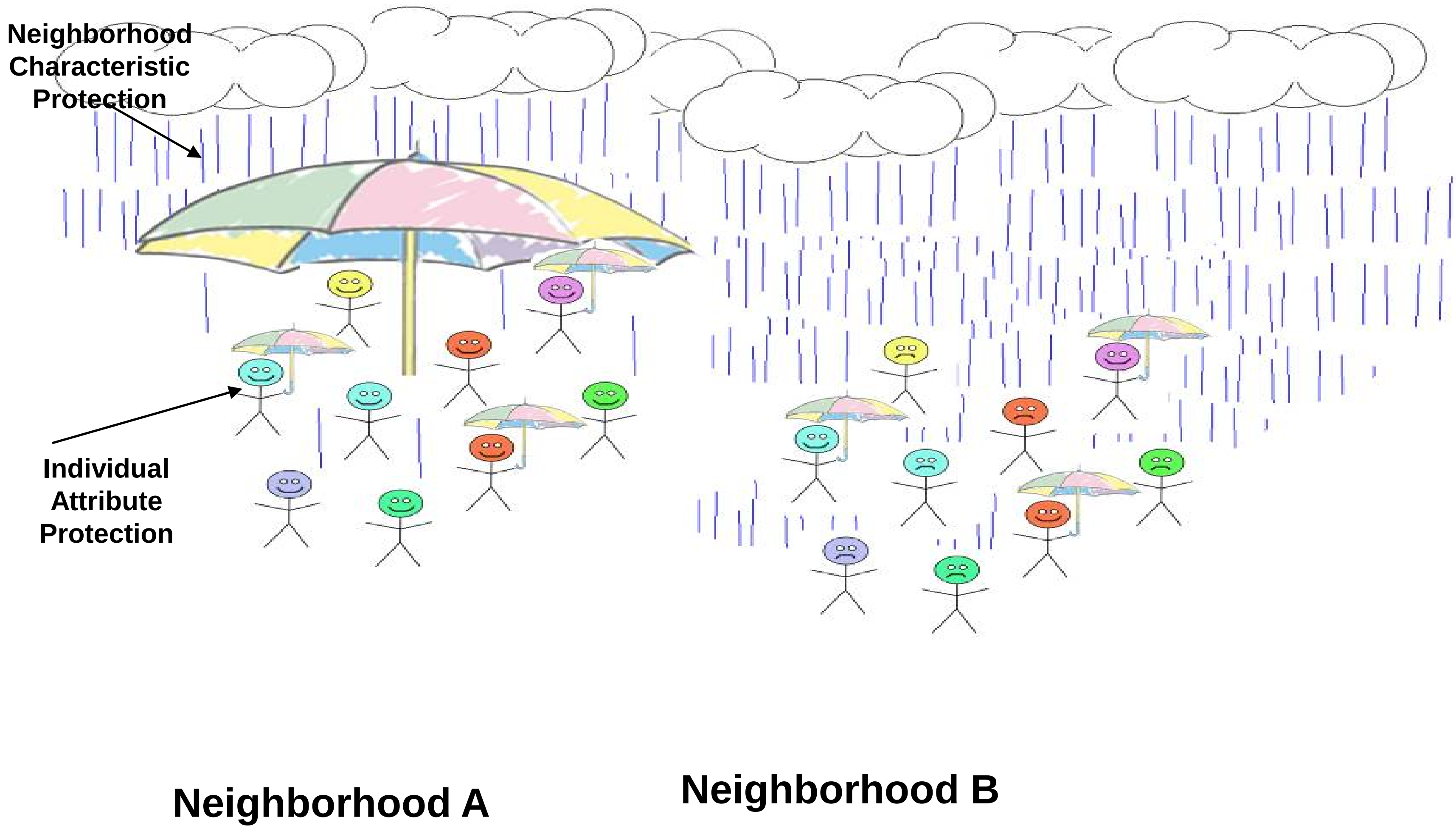


Neighborhood A **Neighborhood B**

Individual and Neighborhood Effects



Individual and Neighborhood Effects



Compositional vs. Contextual

- Compositional – the difference that people make to places

- Contextual – the difference that places make to people

Types of Multilevel Question

Question	Number of levels
Is there an association between neighborhood food availability and diet	
Is there an association between BMI and changes in blood pressure over time?	
Is there an association between provider characteristics and treatment compliance among patients with type 2 diabetes	

Types of Multilevel Question

Question	Number of levels
Is there an association between neighborhood food availability and diet	2 levels: people nested within neighborhoods
Is there an association between BMI and changes in blood pressure over time?	2 levels: time nested within people
Is there an association between provider characteristics and treatment compliance among patients with type 2 diabetes	2 levels: people nested within providers

Types of Multilevel Question

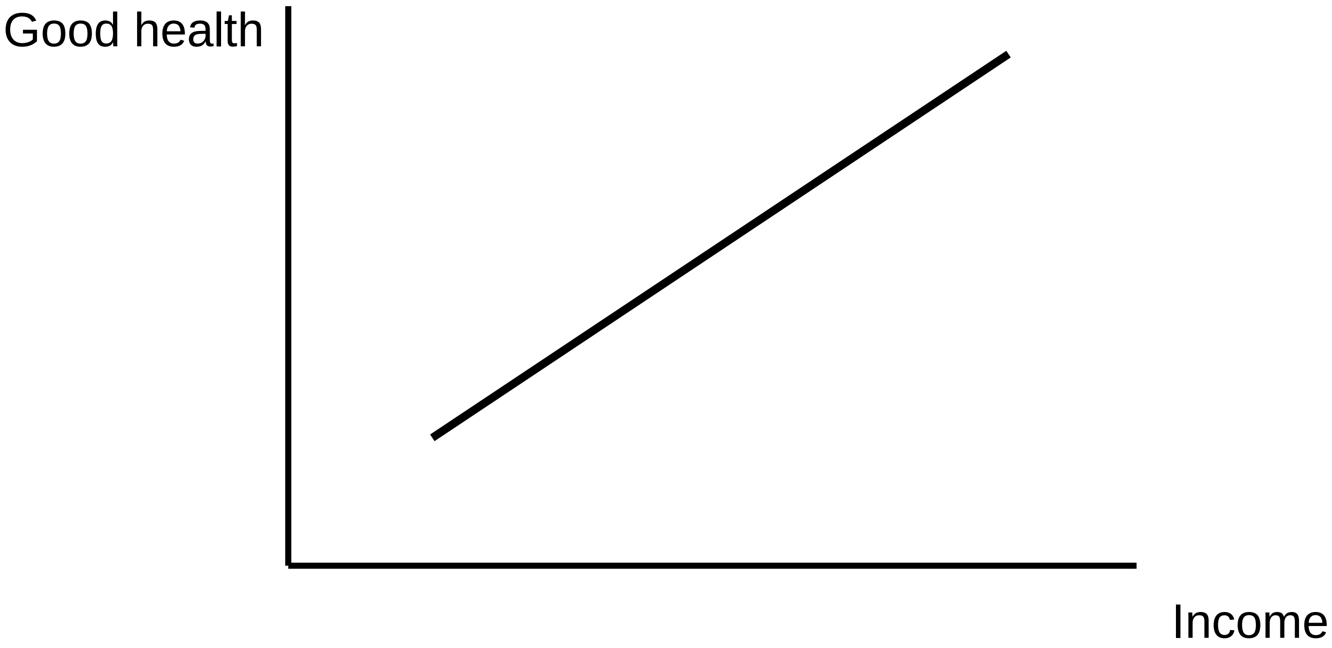
Question	Number of levels
Is there an association between neighborhood food availability and changes in diet over time?	
Is there an association between provider and hospital characteristics on breast cancer recurrence for women with breast cancer?	
Is there an association between city and neighborhood level characteristics in relation changes in blood pressure over time	

Types of Multilevel Question

Question	Number of levels
Is there an association between neighborhood food availability and changes in diet over time?	3 levels: time, individuals, neighborhoods
Is there an association between provider and hospital characteristics on breast cancer recurrence for women with breast cancer?	3 levels: individuals, providers, hospitals
Is there an association between city and neighborhood level characteristics in relation changes in blood pressure over time	4 levels: time, individuals, neighborhoods, cities

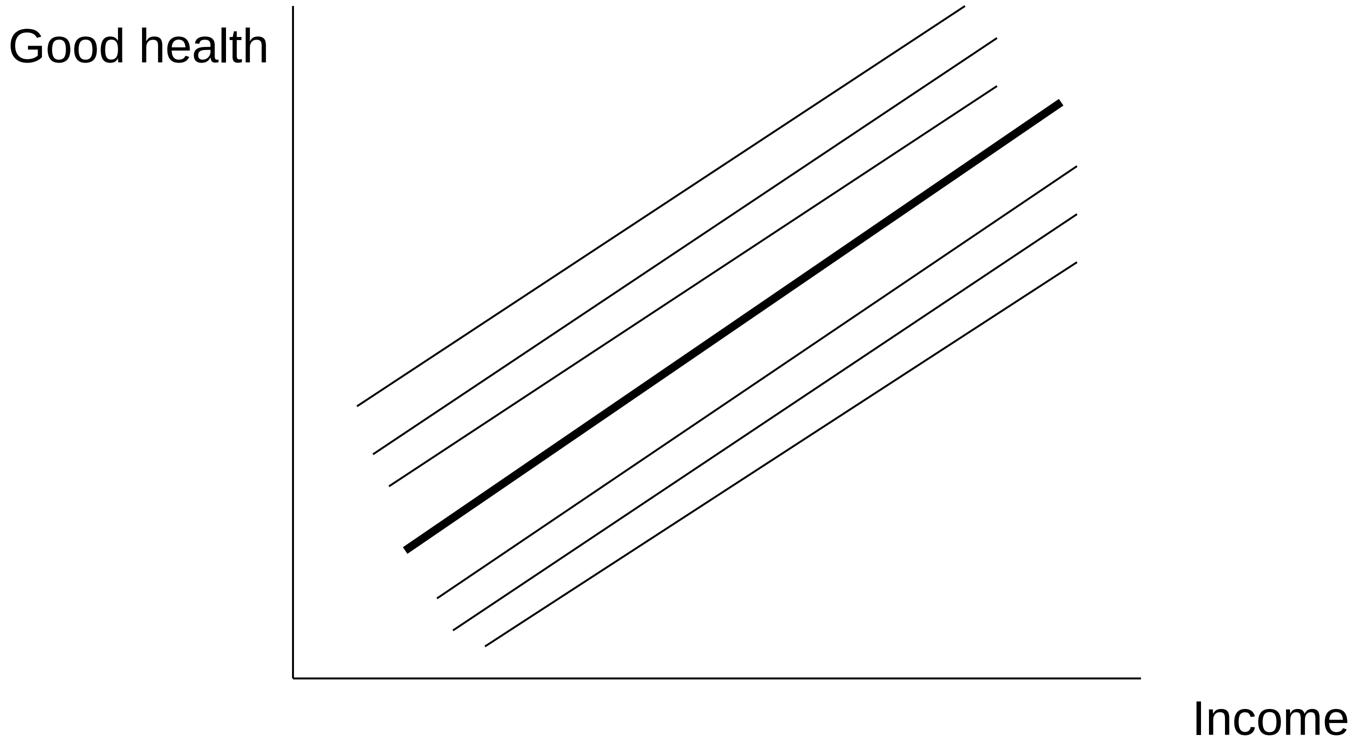
Multi-level Analysis: A Conceptual Overview

Individuals (level 1) nested within six neighborhoods (level 2).



Fixed intercept, Fixed slope Model – Ignoring Neighborhood Context

Random intercepts, Fixed slopes



Each neighborhood represented by a separate line at varying distances from *average* relationship indicated by thick line, i.e. intercepts allowed to vary.

Putting it into equations…

Individual-level regression model

$$y_i = \beta_0 x_{0i} + \beta_1 x_{1i} + (\varepsilon_{0i} x_{0i})$$

Outcome:

Good health score

Fixed part of regression model:

Intercept & slope, where x_1 = income

Random part of regression model:

"Residual"

Neighborhood regression model

(assuming health depends only on neighborhood context)

Micro model (for health of *i*th individual in *j*th neighborhood):

$$y_{ij} = \beta_{0j}x_{0ij} + \varepsilon_{0ij}x_{0ij}$$

Macro model (at neighborhood level, allowing intercept to vary):

$$\beta_{0j} = \beta_0 + u_{0j}$$

Health in each of j neighhorhoods depends on fixed average β0, plus random difference allowed to vary for each neighborhood (uoj).

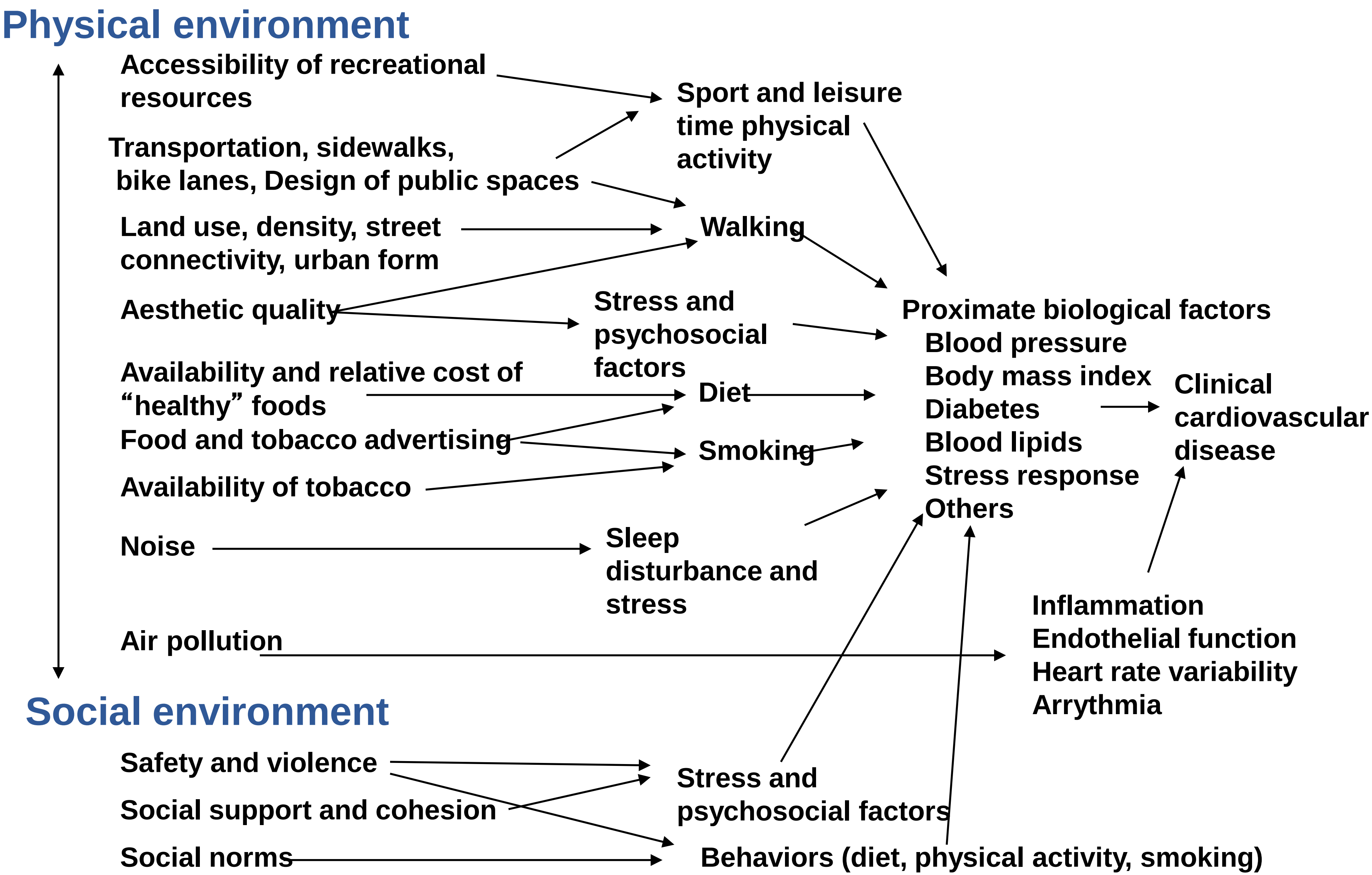
Putting individual and neighborhood models together…

Random intercept model

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + (u_{0j} x_{0ij} + \varepsilon_{0ij} x_{0ij})$$

Intra-class correlation (ICC)

- Partitions variance in multi-level models attributable to individual vs. neighborhood levels
- ρ = (between-neighborhood variance) / (between-neighborhood variance) + (between-individual, within-neighborhood variance)
- Typically ranges between 5-10% in neighborhood studies (depending on health outcome)



Diez Roux AV, 2003: Journal of Urban Health

Multi-Ethnic Study of Atherosclerosis

- A longitudinal study to identify risk factors for sub-clinical atherosclerosis
  - 45-84 years; multi-ethnic
  - Free from clinical CVD at baseline (2000)

MESA Neighborhood Study

Comprehensive Assessment of Neighborhood Context

- Measurement Tools
  - Surveys
  - Geographic Information Systems (GIS)
  - Systematic social observation

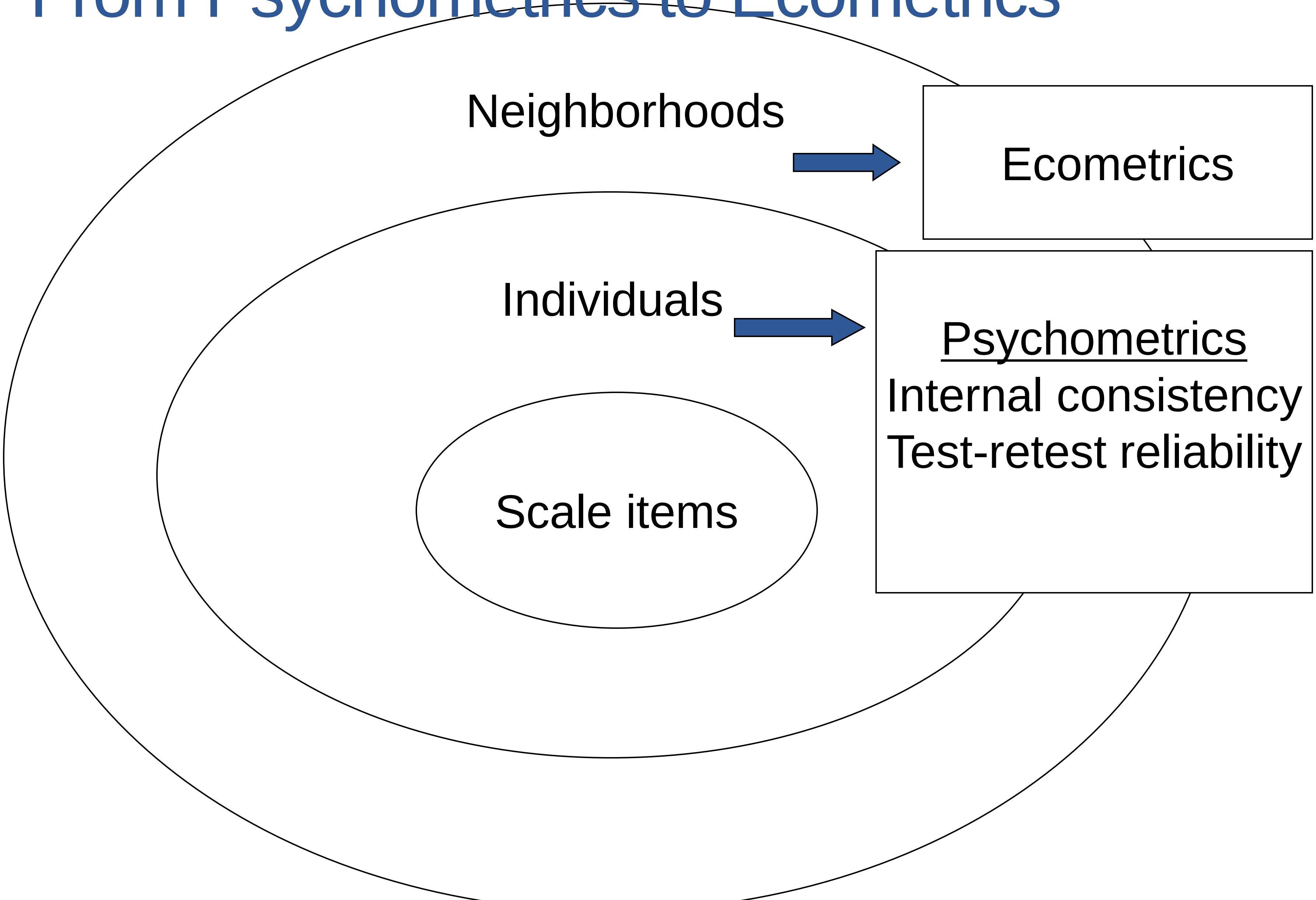
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
MESA Exams	EXAM1		EXAM2		EXAM3	EXAM4					EXAM5		
MESA Neighborhood	EXAM1			NEIGHBORHOOD							EXAM5		
Community Survey					1 (NY, MD, NC)		2 (CA, NY)					3 (ALL SITES)	

GIS Measures: National Industries and Economies Data

Development of Survey-based Measures

Walking/PA environment	1. My neighborhood offers many opportunities to be physically active. 2. Local sports clubs and other facilities in my neighborhood offer many opportunities to get exercise. 3. It is pleasant to walk in my neighborhood. 4. The trees in my neighborhood provide enough shade. 5. In my neighborhood it is easy to walk places. 6. I often see other people walking in my neighborhood. 7. I often see other people exercise (for example jog, bicycle, play sports) in my neighborhood.
Availability of healthy foods	1. A large selection of fresh fruits and vegetables is available in my neighborhood. 2. The fresh fruits and vegetables in my neighborhood are of high quality. 3. A large selection of low fat products is available in my neighborhood.
Safety	1. I feel safe walking in my neighborhood day or night. 2. Violence is not a problem in my neighborhood. 3. My neighborhood is safe from crime
Social cohesion	1. People around here are willing to help their neighbors. 2. People in my neighborhood generally get along with each other. 3. People in my neighborhood can be trusted. 4. People in my neighborhood share the same values

From Psychometrics to Ecometrics



Raudenbush SW, Sampson RJ (1999)

From Psychometrics to Ecometrics

$$\text{Level 3}: \beta_{00k} = \gamma_{000} + U_{00k} \rightarrow U_{00k} \sim N(0, \tau_\eta)$$

$$\text{Level 2}: b_{0jk} = \beta_{00k} + \alpha_{0jk} \rightarrow \alpha_{0jk} \sim N(0, \tau_b)$$

$$\text{Level 1}: Y_{ijk} = b_{0jk} + e_{ijk} \rightarrow e_{ijk} \sim N(0, \sigma^2)$$

Variance components

Raudenbush SW, Sampson RJ (1999)

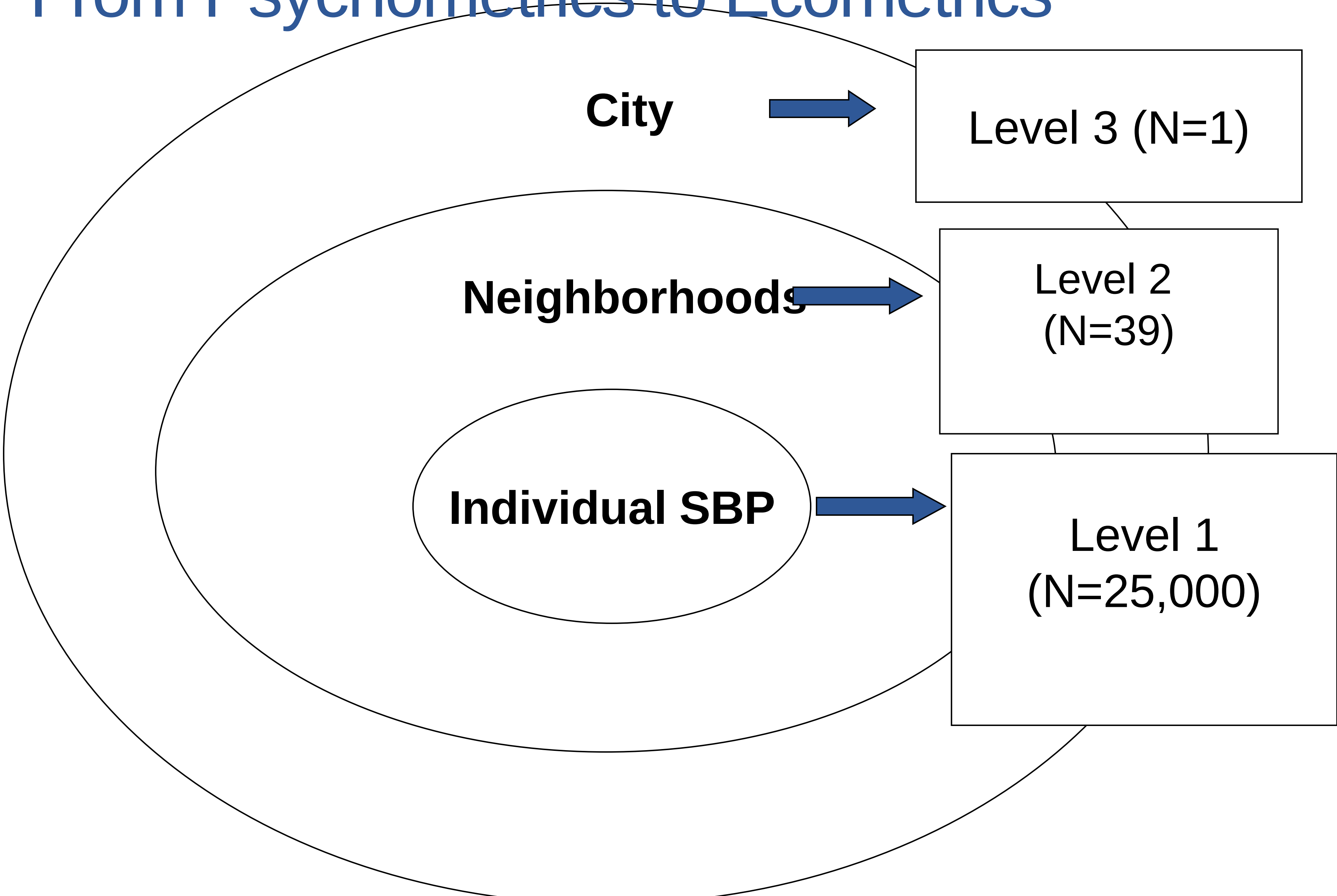
Ecometric Measures

1. Intraneighborhood agreement (ICC)
   - % of variability in neighborhood scale items that lies between neighborhoods as opposed to within neighborhoods
   - range: (0-1)
2. Neighborhood reliability
   - The accuracy of the model estimate as a measures of the true neighborhood dimension
   - range:(0-1)

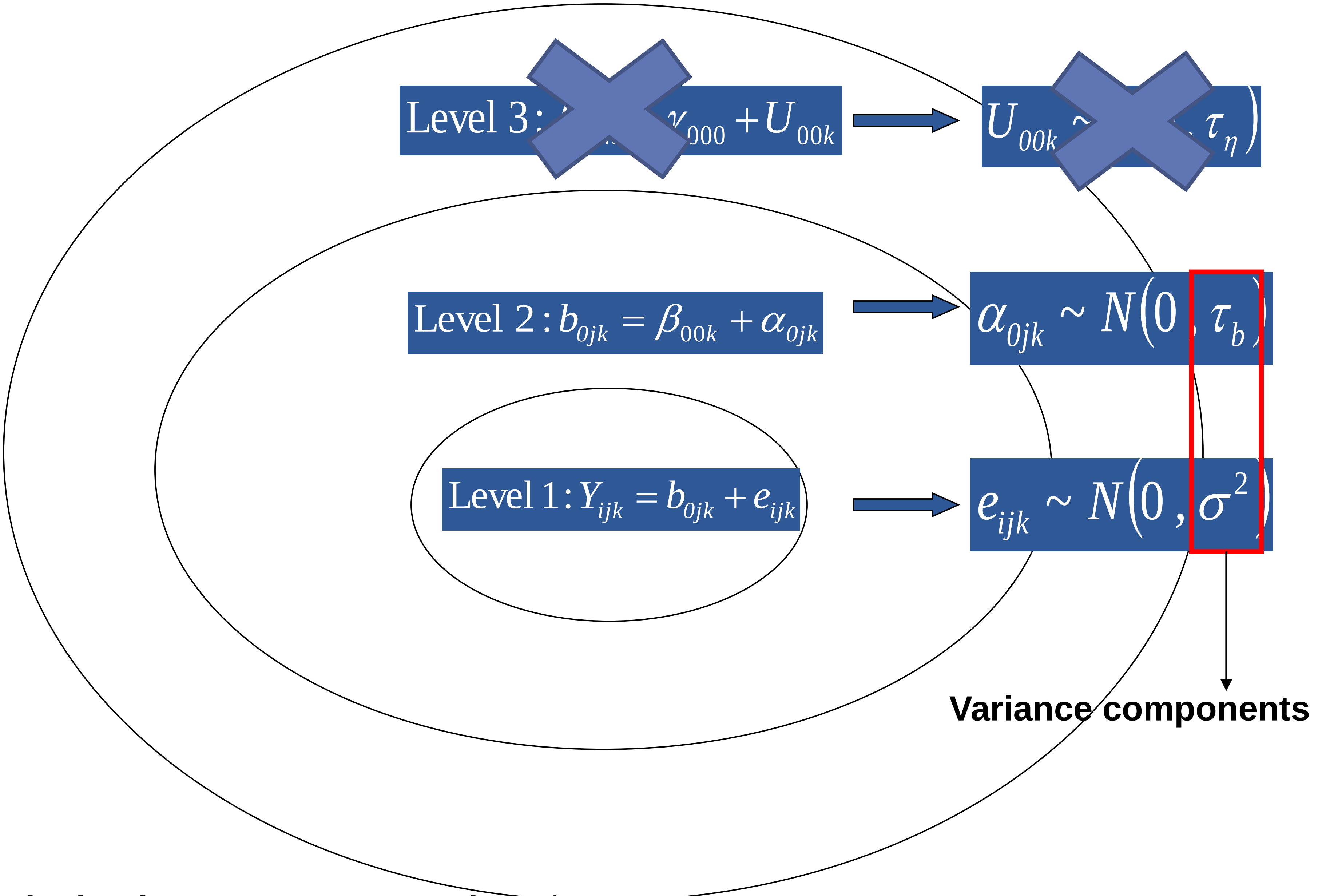
Scales Have Good Psychometric and Ecometric Properties

Construct	Cronbach's Alpha	Test-retest [95% C.I]	Intra-neighborhood Correlation	Neighborhood Reliability
Aesthetic quality	0.75	.83 [.77, .88]	0.51	0.78
Walking environment	0.73	.60 [.47, .71]	0.43	0.73
Availability of healthy foods	0.78	.69 [.57, .77]	0.28	0.64
Safety	0.77	.88 [.83, .91]	0.37	0.77
Violent crime	0.83	.72 [.62, .80]	0.34	0.72
Social cohesion	0.74	.65 [.53, .74]	0.31	0.68

From Psychometrics to Ecometrics



Raudenbush SW, Sampson RJ (1999)



Raudenbush SW, Sampson RJ (1999)

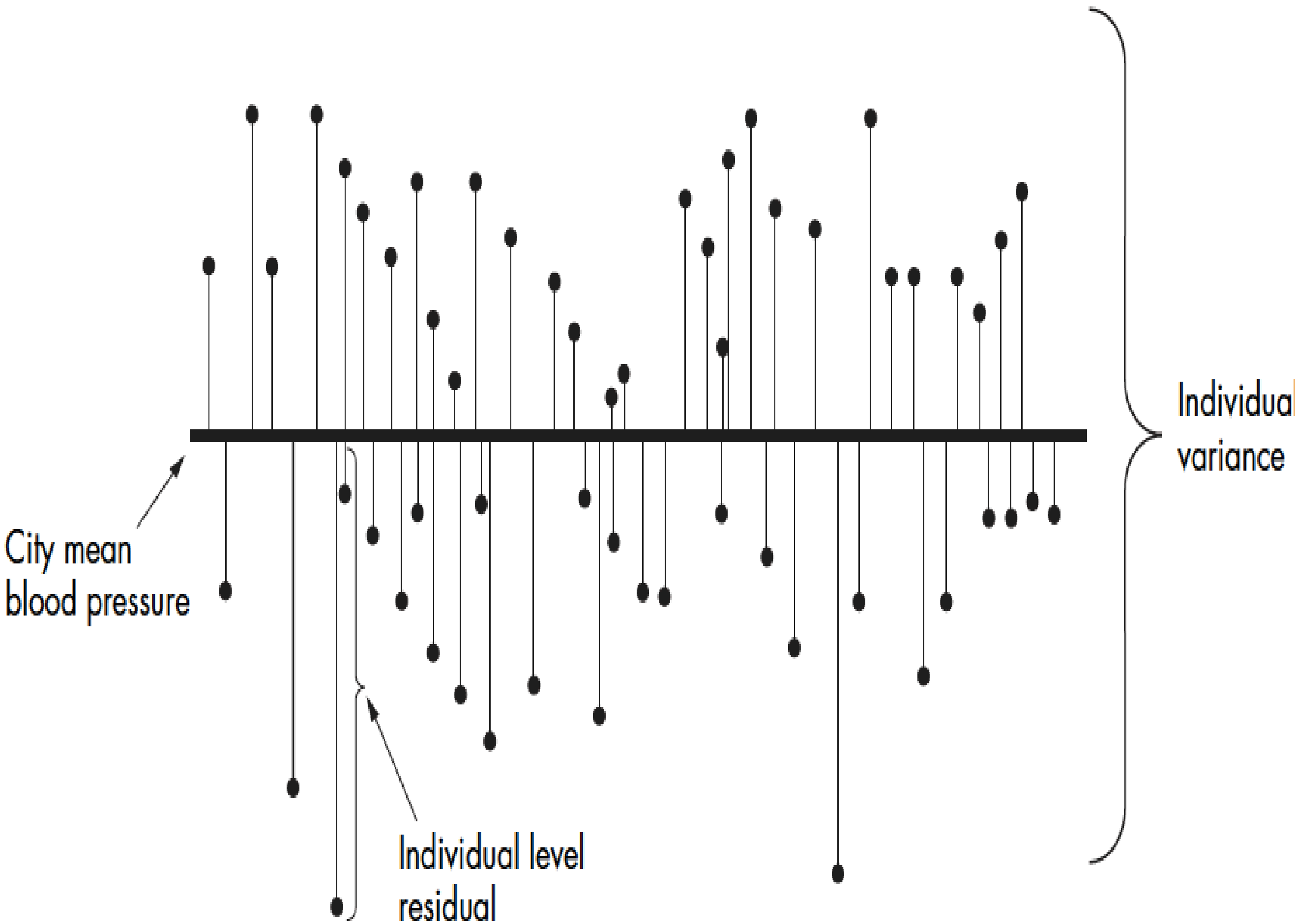


Figure 3 Single level individual information. This figure represents the distribution of individual SBP in the population of the city when we have only single level individual based information. The fact that people are grouped within neighbourhoods is neglected, as we only have individual level data. In this figure the length of the thin vertical line from the black spot to the thick horizontal line represents the individual differences in blood pressure compared with whole city mean (the individual level residuals). The individual variance in single level individual studies is an average summary of these differences. In single level individual analysis we consider all information as if it were at the individual level neglecting possible neighbourhood components.

Merlo J, et al, 2005, J Epidemiol Community Health

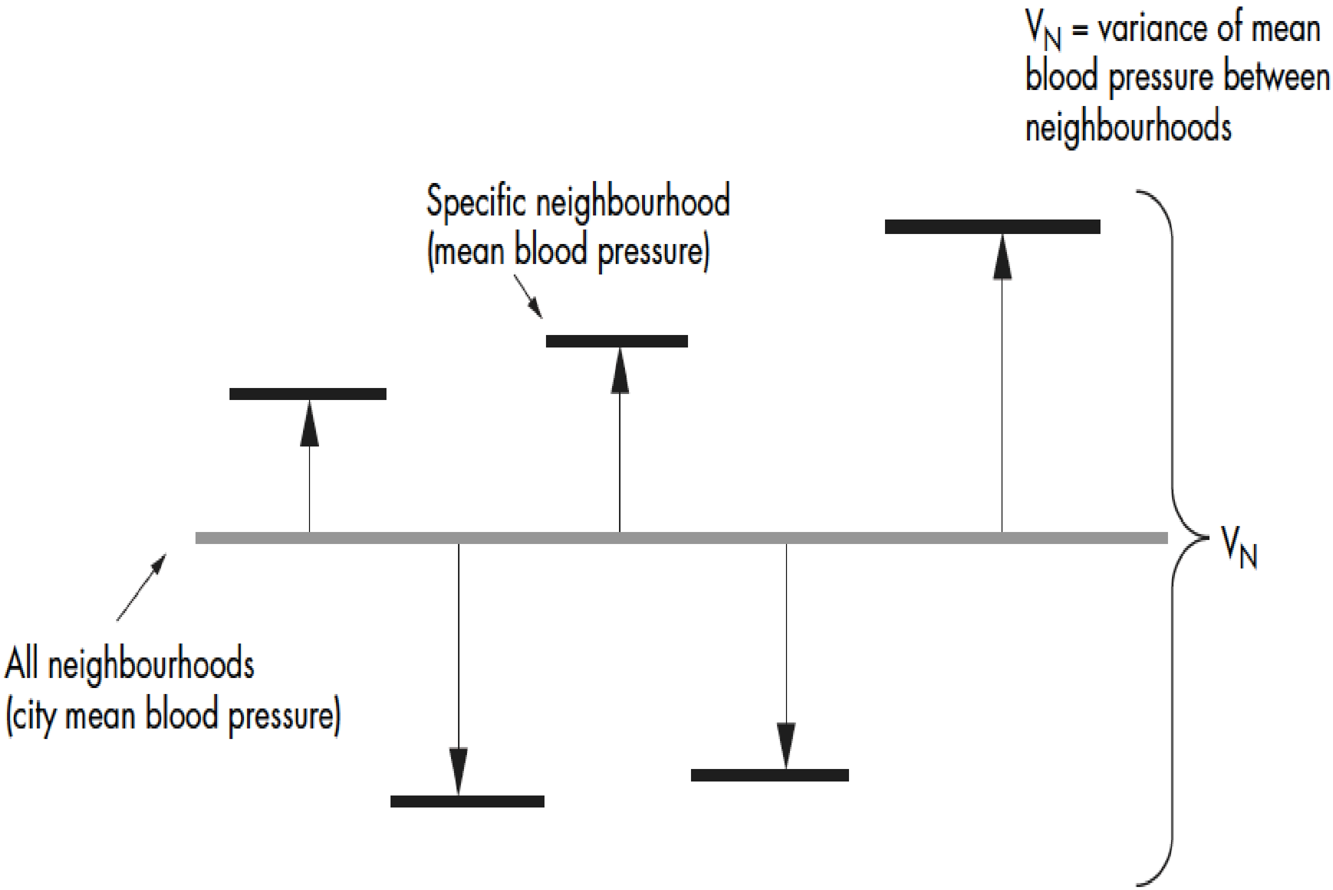


Figure 4 Single level ecological information. In this figure all individual SBP values are aggregated at the neighbourhood level to obtain the neighbourhood mean. We can distinguish differences between the mean blood pressure of each neighbourhood and the mean blood pressure of the whole city (the neighbourhood residuals). These residuals are represented by thick black horizontal lines at the top of a pillow. The neighbourhood variance is a summary of the differences between neighbourhoods. We are unable to observe differences between people (variation in blood pressure within neighbourhoods). In single level ecological analysis we consider all information as if it were at the neighbourhood level neglecting individual components.

Merlo J, et al, 2005, J Epidemiol Community Health

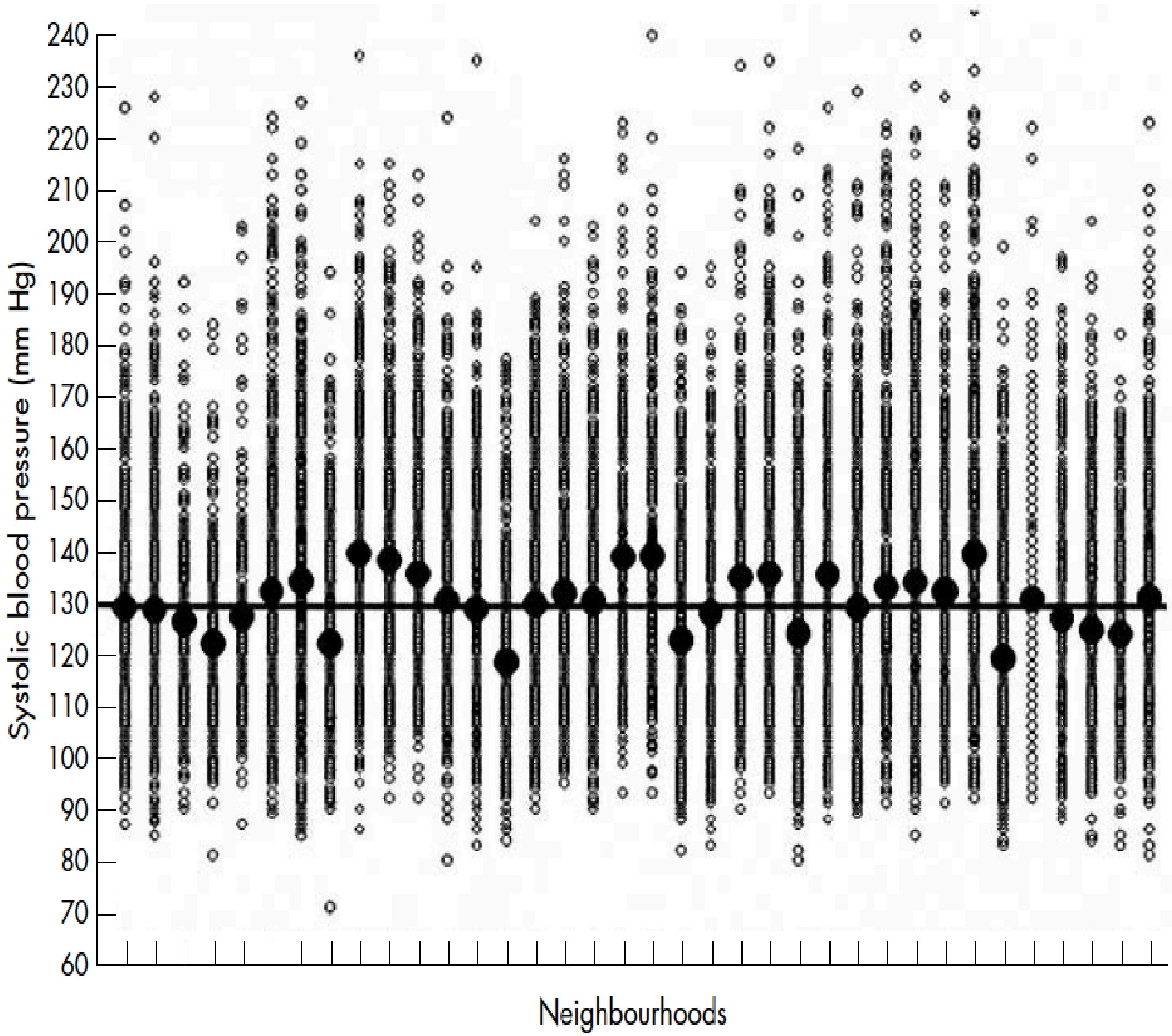


Figure 2 The figure shows the actual SBP values used in our example. The large black dots represent the neighbourhood means. The small circles represent the individual SBP values within neighbourhoods. The horizontal black line represents the city SBP mean.

Merlo J, et al, 2005, J Epidemiol Community Health

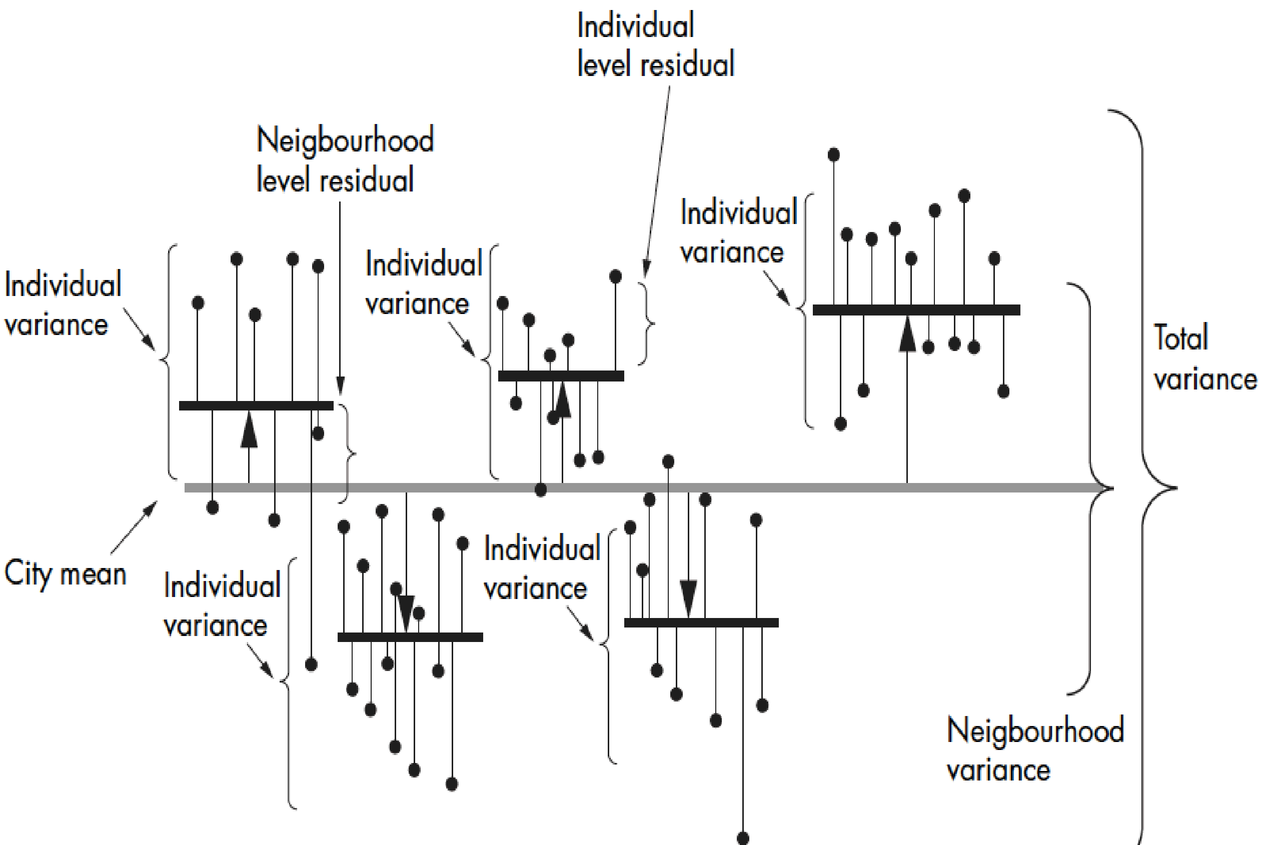


Figure 1 Multilevel information. In this figure the neighbourhood residuals are represented by the length of the pillows between the city SBP mean, represented by a grey colour, and the neighbourhood SBP means represented by thick black horizontal lines. The individual residuals are represented by the length of the vertical lines between the neighbourhood means and the individual SBP values represented by black circles at the top of thin lines. In this figure we do not have any explanatory variable (that is, this figure corresponds to an "empty" model) as we are only interested in analysing how individual blood pressure differences are partitioned in a variability that exists between people from the same neighbourhood and a variability that exists between neighbourhoods. In this figure we can imagine that the neighbourhood means (short thick lines) pull up or pull down all the individual SBP values belonging to the same neighbourhood, even if individual level variability remains within neighbourhoods. The mathematical expression of the intraclass correlation can be visually understood in figure 1. Figure 1 is a graphic combination of figures 3 and 4.

Merlo J, et al, 2005, J Epidemiol Community Health

Table 1 Multilevel, individual and ecological linear regression analysis of systolic blood pressure (SBP) in 25 000 people living in the 39 neighbourhoods (hypothetical data)

	Multilevel regression analysis (empty model)	Single level individual analysis*	Single level ecological analysis†
Fixed effects			
Mean SBP of the city	130.2	130.2	130.0
Random effects			
Multilevel measures of health variation			
Variance (SE)			
Between neighbourhoods	36.2	–	37.0
Between individuals	433.4	468.1	
Intraclass correlation (ICC)	ICC = 0.08	–	–
Deviance (goodness of fit of the model)‡	222764	224532	–

"Fixed effects" and "random effects" are expressions that are often used in MLRA. In very simple terms, fixed effects are used to model means whereas random effects are used to model variance. *In the individual level analysis we ignore the existence of the neighbourhood level. †In the single level ecological analysis we aggregate the individual level information on SBP to the neighbourhood level, neglecting the existence of the individual level. ‡The deviance expresses the goodness of fit of the model, and was calculated using the $-2\times$log likelihood (IGLS). The Bayesian deviance information criterion (DIC) gave similar results. The multilevel models fit the data much better than the single level individual analysis. The ecological analysis is not comparable as it is based on 39 neighbourhoods rather than on 25 000 people. We refer elsewhere for an extended explanation of the deviance statistic.[44 45]

Merlo J, et al, 2005, J Epidemiol Community Health

Example: Two-level Patients and Hospital Characteristics

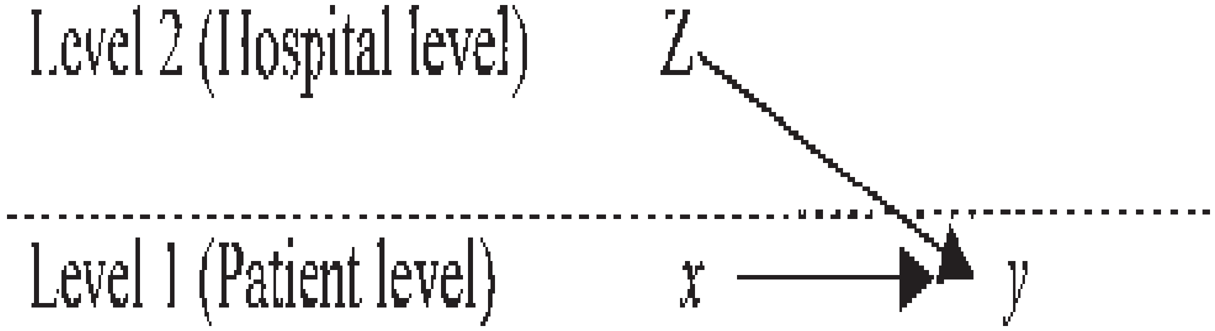


FIGURE 1. The structure of a multilevel proposition. x = patient characteristics (age, type of admission, number of diagnoses); y = probability of developing pneumonia; Z = hospital characteristics (nurse staffing, ownership).

Cho SH, Nursing Research 2003

Example: Two-level Patients and Hospital Characteristics

TABLE 1. Estimates for Multilevel Logistic Regression of Pneumonia as a Function of Patient and Hospital Characteristics

	Model 1		Model 2		Model 3	
Fixed Effect	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
γ_{00} = intercept	−3.1194*	.1173	−2.9275*	.1369	−2.8776*	.5366
γ_{10} = coefficient of AGE			0.0291*	.0052	0.0292*	.0052
γ_{20} = coefficient of ADMIT			−0.9765*	.1660	−0.9774*	.1676
γ_{30} = coefficient of NDX			0.0543*	.0204	0.0546*	.0205
γ_{01} = coefficient of OWN					0.1578	.2886
γ_{02} = coefficient of STAFF					−0.0146	.0873
Random Effect	Variance Component	Standard Deviation	Variance Component	Standard Deviation	Variance Component	Standard Deviation
τ_{00} = var(U_{0j}) intercept variance	.3551*	.5959	.3792*	.6158	.3747*	.6121
Deviance	11,705.93		11,592.17		11,591.67	
Number of parameters	2		5		7	

Note. AGE = patient age; ADMIT = type of admission; NDX = number of diagnoses at admission; OWN = hospital ownership; STAFF = nurse staffing.
*$p < .05$

Cho SH, Nursing Research 2003

Oakes' Critique of Multilevel Approach

- Social stratification sorts individuals into different neighborhoods

- By controlling for individual compositional effects of social class, the multilevel analyst runs the risk of making adjustments until "there is nothing for the neighborhood variables to explain"

- But if we don't control for individual social class, we run the risk of residual confounding

- Even if we could find poor people living in affluent communities (or vice versa), "these people are exceptions to the rule and should not be given the same level of statistical credence as the majority"

Oakes JM. *Soc Sci Med* 2004; 58:1929-52.

Dealing with Selection

- Standard Approach: Control for a set of measured individual characteristics
- Pitfall #1: Undercontrol
  - Variables involved in the sorting of people into neighborhoods are omitted or poorly measured
- Pitfall #2: Overcontrol
  - Including in the model variables that are causally "downstream" from W_j
  - Especially those on the causal pathway between W_j and Y_{ij}

Summary

- There is no right answer on model building in health disparities research
- Decisions depend on the research question and the conceptual framework guiding the research
- Analytic challenges are not unique to health disparities research, but often this research is more scrutinized in the literature