Systematic Measurement Error's Influence on Estimating and Understanding Health Disparities

Adam C. Carle, M.A., Ph.D.
adam.carle@cchmc.org
James M. Anderson Center for Health Systems Excellence
Cincinnati Children's Hospital Medical Center
University of Cincinnati School of Medicine
University of Cincinnati College of Arts and Sciences

Introduction

- Population health research seeks to determine the "health outcomes of a group of individuals, including the distribution of such outcomes within the group."
    - (Kindig & Stoddart, 2003).

- Populations can reflect geographic regions and/or socially defined groups.
    - – e.g., Different racial and ethnic groups.

Introduction

- Approach "requires" measures of health outcomes of populations.
  - Urgent need for reliable and valid measures.

- Approach also allows focus on disparities across subpopulations.

Introduction

- Health disparities.
  - AHRQ defines disparities as inequalities in health or health care that one population experiences relative to another.
    - (AHRQ, 2010).
  - IOM defines disparities as racial or ethnic differences in quality not due to access-related factors or clinical needs, preferences, and intervention appropriateness.
    - (IOM, 2002).

Cincinnati Children's

Introduction

- Others highlight distinction between:
  - Inequalities.
    - Differences.

  - Inequities.
    - Avoidable and unfair health inequities.
      - (Asada, 2005).

- All highlight differences in distribution of health outcome(s) across population subgroups.

Introduction

- Accurately understanding differences in distribution of an outcome across heterogeneous populations requires equivalent measurement across population.
  - (Stahl & Hahn, 2006).

- Little research addresses possibility that systematic measurement error influences population health research.

Introduction

- Before making cross-group comparisons, *must* consider measurement equivalence.
- Do *observed* differences reflect *true* differences?
- Or, do differences result from systematic measurement error?

Measurement Bias

- Refers to possibility that individuals with identical health respond dissimilarly to questions about their health as a function of their race or ethnicity.
  - (Mellenbergh, 1989).

- Individuals with identical health statuses from different backgrounds may respond differently to questions about their health.
  - Should respond similarly, but don't.

- Systematic measurement error.
  - Measurement bias.
  - Differential item functioning (DIF).

Cincinnati Children's

Measurement Bias

- Measurement bias:
  - Individuals identical on measured construct respond dissimilarly as a function of group membership.
    - e.g., White, Black, Hispanic.

- Measurement equivalence:
  - Denotes equal endorsement probabilities for individuals with equal construct values.
  - Group membership does not predict differences.

Why Study Bias?

- Generally decreases reliability and validity.
    - (Knight & Hill, 1998).
- Attenuate or accentuate group differences.
    - (Carle, 2008).
- Lead to inaccurate diagnoses.
    - (Carle, 2009).
- Can render cross group comparisons impossible.
    - (Prelow, et al., 2002).

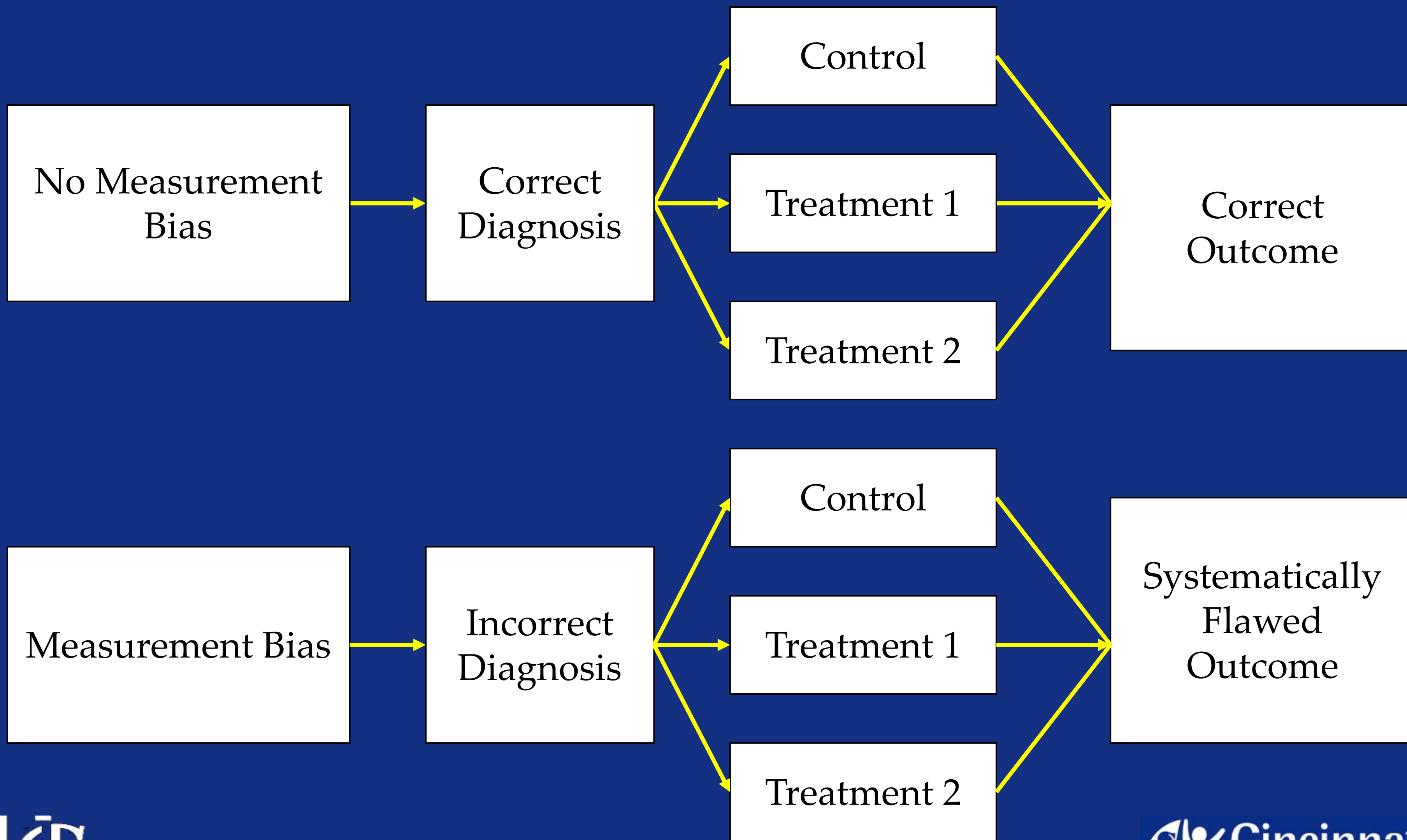
Why Study Bias?

- Without establishing equivalent measurement across the heterogeneous population, field cannot:
  - Comparatively evaluate what works best for whom.
  - Draw strong conclusions about disparate outcomes.
  - Support evidence-based practice and policy.
  - Address health disparities.

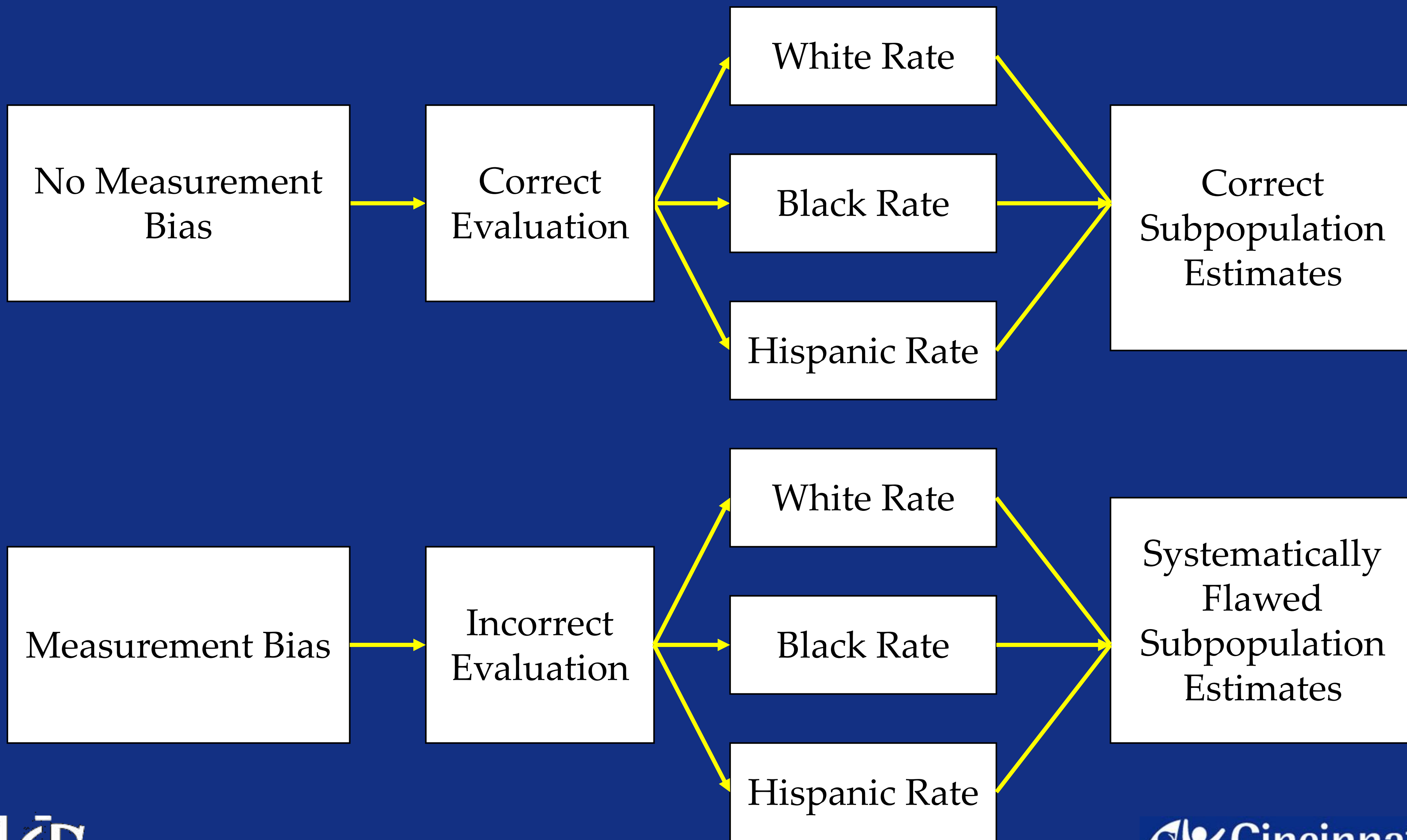
- How might this influence research?

Cincinnati Children's

Why Study Bias?



Why Study Bias?



Evaluating Bias

- Latent variable models potently investigate bias.
    - (Millsap & Kwok, 2004, Muthén, 1989).
- Equations describe the relations among item set.
- Examine the cross-group equivalence of the measurement parameters in the equations.
    - (Millsap & Yun-Tien, 2004).
- Differences in these parameters across groups reflect bias.

Evaluating Bias

- Multiple group (MG) confirmatory factor analyses for ordered-categorical measures (CFA-OCM).
  - One popular method.
  - Accounts for categorical nature of data.

Evaluating Bias: MG-CFA-OCM

- Let X_{ij} equal the *i*th individual's score on the *j*th ordered-categorical item.
  - Let the number of items be *p* (*j* = 1, 2, .., p).
  - Scores, *m*, range {0, 1, …, *s*}.

- We assume a continuous latent response variate, X_{ij}^{*}, determines observed responses.

Evaluating Bias: MG-CFA-OCM

- A threshold value on X_{ij}^* determines responses:
  - If X_{ij}^* less than the threshold, respond in one category.
  - If X_{ij}^* greater than threshold, respond in at least next highest category.

$$X_{ij} = m \quad \text{if} \quad \nu_{jm} \leq X_{ij}^* \leq \nu_{j(m+1)}$$

- $\{\nu_{j0}, \nu_{j1}, \ldots, \nu_{j(s+1)}\}$ represent threshold parameters.

Evaluating Bias: MG-CFA-OCM

- Suppose some factor or set of factors, ξ, is responsible for the observed scores.

- X_{ij}^{*} relates to the factor(s) as follows:

$$X_{ij}^{*} = \tau_j + \lambda_j'\xi_i + \varepsilon_{ij}$$

Evaluating Bias: MG-CFA-OCM

$$X_{ij}^{*} = \tau_j + \lambda_j' \xi_i + \varepsilon_{ij}$$

- τ_j : Latent intercept parameters.
  - Similar to intercepts in regression.

- λ_j' : Factor loadings.
  - Similar to correlations.
  - Represents how strongly the latent response variate relates to the factor(s).

Evaluating Bias: MG-CFA-OCM

$$X_{ij}^{*} = \tau_j + \lambda_j' \xi_i + \varepsilon_{ij}$$

- ξ_i : Individual's level of the latent trait(s).
- ε_{ij} : Variance not attributable to the factor(s).
  - Includes measurement error.

Evaluating Bias: MG-CFA-OCM

- Subscript parameters to allow group differences.

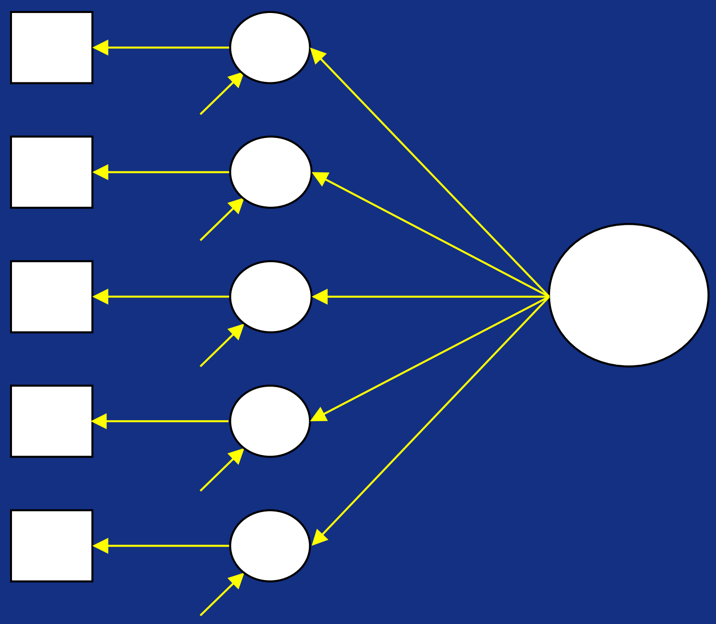
- Begin with the least restricted cross-group model.
  - Successively constrain subsequent models.
  - Model suitability addressed through goodness-of-fit-indices (GFIs).
    - (Hu & Bentler, 1999).

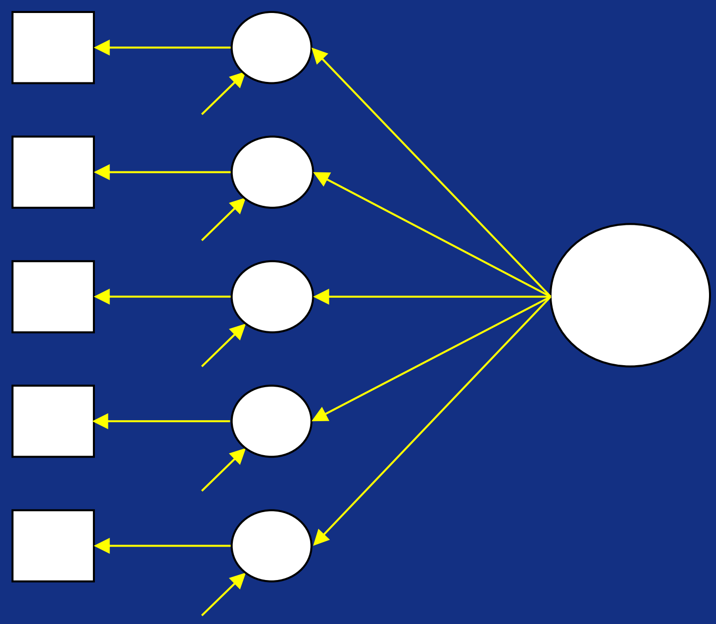
- If GFI set suggests fit not tenable at a given step, bias exists.
  - Bias in at least one statistical parameter.
  - Cross group comparisons not appropriate without adjustment.

?


"A BUSHEL IS A UNIT OF WEIGHT EQUAL TO FOUR PECKS."
WHAT'S A PECK?
A QUICK SMOOCH.
YOU KNOW, I DON'T UNDERSTAND MATH AT ALL.

Evaluating Bias: MG-CFA-OCM





Evaluating Bias

- Methodological and substantive issues can limit MG-CFA-OCM.

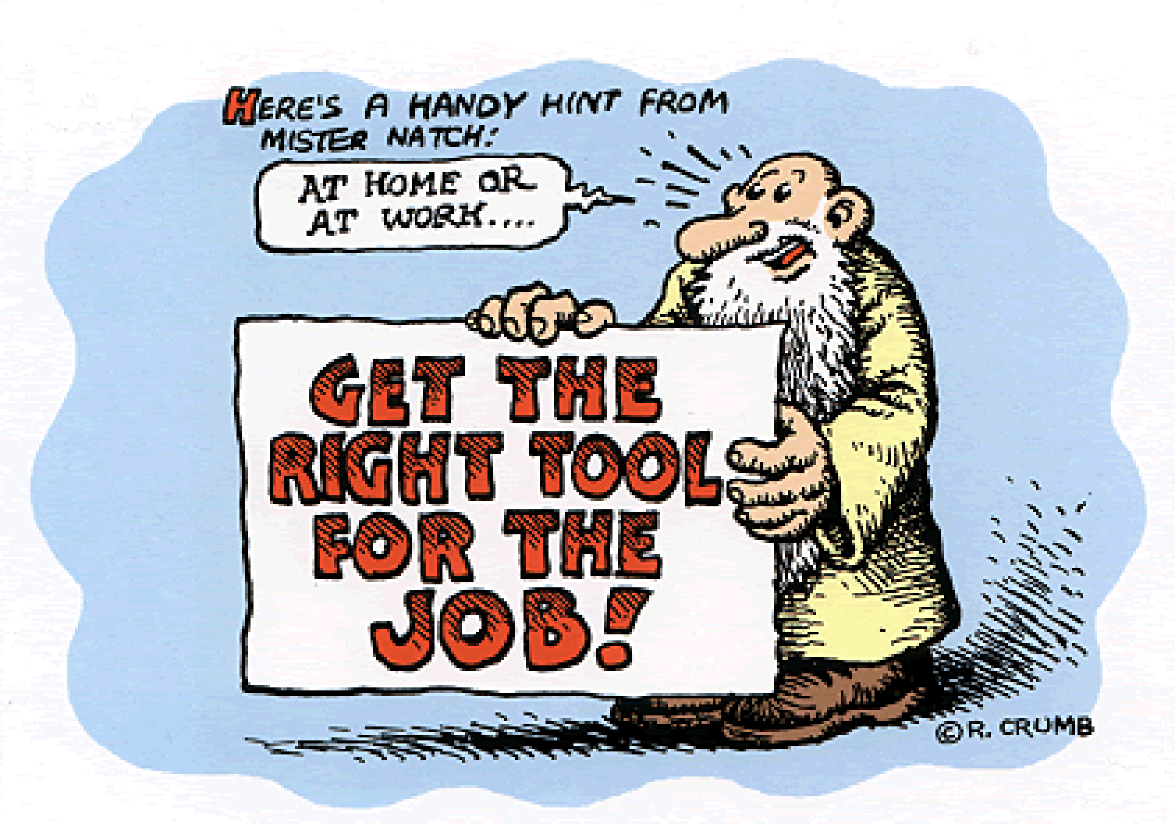
- Difficult to incorporate multiple grouping variables simultaneously.

- Why does this matter?

- Bias may result from other variables that covary with ethnicity.
  - Educational attainment.
  - Income/poverty status.

Observational Research

- Difficult to simultaneously include multiple variables in “traditional” latent variable approaches.
- Failure to include available information in model estimation may lead to erroneous conclusions.
- What do we do?


HERE'S A HANDY HINT FROM MISTER NATCH:
AT HOME OR AT WORK....
GET THE RIGHT TOOL FOR THE JOB!
©R. CRUMB

MG-MIMIC Models

- Multiple group (MG) multiple indicator, multiple cause (MIMIC) models.

  - Build on developments in structural equation modeling, IRT, and CFA-OCM.
    - (Jones, 2003; Jones, 2006; Muthén, 1989)

- Control for "extra" variables by incorporating them as covariates.

MG-MIMIC Models

- Simultaneously:

- Examine and control response differences due to covariates (e.g., SES)…..

  – And

- Allow bias investigation across groups with background variable effects removed.

- More fully address heterogeneity within and across groups.

Evaluating Bias: MG-MIMIC

$$X_{ij}^{*} = \tau_j + \lambda_j' \xi_i + \kappa_i x_i + \varepsilon_{ij}$$

- x_i represents the covariate.
- Parameters in κ capture the direct effect of the covariate on question responses.
  - Addresses whether covariate influences measurement.

MG-MIMIC Models

- But, covariate may predict values of the measured trait.
  - e.g., Education may predict mental heath symptomatology.
- As a result, covariate may indirectly influence measurement.
- A structural component to the model captures these notions.

Cincinnati Children's

Evaluating Bias: MG-MIMIC

$$\xi_i = \alpha + \gamma x_i + \zeta$$

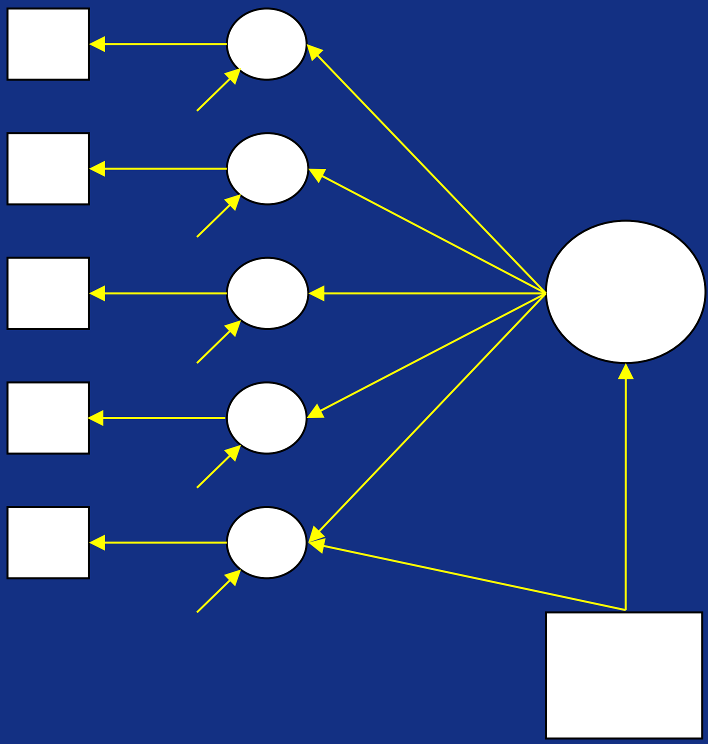
- x_i represents the covariate.
- Parameters in γ capture the indirect effects.
- α represents the average value of the factor.
- ζ correspond to the residuals in the model.

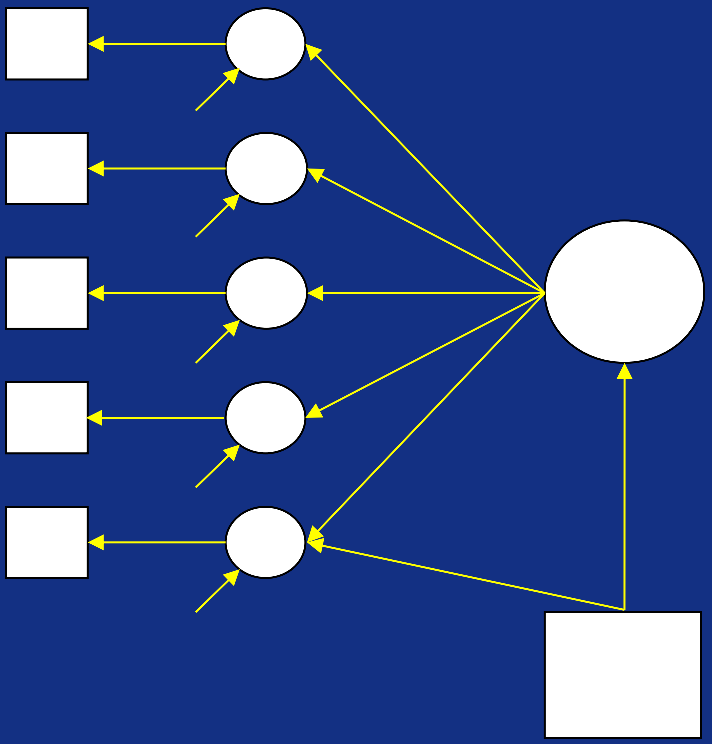
Evaluating Bias: MG CFA-OCM

- Subscript parameters to allow group differences.

- To the extent that cross-group constraints in κ_g, τ_g, Λ_g, $\left\{\nu_{jg0}, \nu_{jg1}, \ldots, \nu_{jg(s+1)}\right\}$, and Θ_g lead to problematic GFIs, measurement bias presents.

MG-MIMIC Models





Using MG-MIMIC to assess Bias

- How do we do this in practice?

- Use series of hierarchically nested models.

- Examine tenability of cross-group constraints in the measurement parameters.
    - (Muthén, 1989; Jones, 2006; Millsap & Yun-Tien, 2004).

Using MG-MIMIC to assess Bias

- Begin with the least restricted cross-group model.
- Successively add cross-group equivalence constraints in subsequent models.
- Bias assessed in each set of measurement parameters separately.
- Model suitability addressed through several goodness-of-fit-indices (GFIs).
    - (Hu & Bentler, 1999).

Using MG-MIMIC to assess Bias

- If GFI set suggests fit not tenable at a given step, bias presents.
  - Bias in at least one statistical parameter.
  - Cross group comparisons not appropriate without adjustment.

- If GFIs suggest tenable model fit, analyses examine equivalence constraints in next parameter set of interest.

Current Study

- Utilized MG-MIMIC to examine alcohol abuse behavior.
  - Probed for bias across race and ethnicity in the 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions.
    - (NESARC: Grant, Kaplan, Shepard, & Moore, 2003).
- MG-MIMIC models simultaneously included participant's education and income level in analyses.
  - Education: No high school vs. high school or more.
  - Income: Below and above 200% poverty level.

Cincinnati Children's

Methods

- Participants (n = 25,512).
  - White: n = 16,480.
  - Black: n = 4,139.
  - Hispanic: n = 4,893.

- Represent noninstitutionalized US adults.

- Ten (10) questions operationalized DSM-IV construct of alcohol abuse.

- All analyses used Mplus and incorporated complex design and weights.
    - (5.1: Muthén & Muthén, 1998-2007).

Methods

- Complex, multistage sampling approach used stratified random sampling.
  - Oversampling insured increased accuracy for Hispanics, African-Americans, young adults.

- Design weights adjust for varying selection probability, other issues, and make data nationally representative.

  - Used zero weight approach to subset data.
    - (Korn & Graubard, 2003).

Results: MG-MIMIC

- Similar unconstrained model fit across all grouping variables.
  - RMSEA = 0.021
  - CFI = 0.98
  - TLI = 0.98
  - McDonald's NCI > 0.99
  - Gamma Hat = 0.999

- Items should measure a similar construct and have comparable meaning regardless of race, ethnicity, education, or poverty status.

Results: MG-MIMIC

- Model examining direct effects of poverty and education uncovered bias.

  - $\Delta\chi^2 = 191.19$ (24, $n = 25{,}512$, $p < 0.01$)

- Univariate analyses identified problematic constraints.

Results: MG-MIMIC

- For more highly education whites:
  - Easier to endorse alcohol:
    - Caused trouble with job/school ($\Delta\kappa$ = - 0.05).
    - Caused trouble with family/friends ($\Delta\kappa$ = - 0.044).
    - Led to fights ($\Delta\kappa$ = -0.038)
    - Led to legal problems ($\Delta\kappa$ = - 0.035)
  - More difficult to endorse:
    - Driving while drinking ($\Delta\kappa$ = 0.044)
    - Driving after drinking ($\Delta\kappa$ = 0.024)
    - Harmful situations while drinking ($\Delta\kappa$ = 0.034)

Cincinnati Children's

Results: MG-MIMIC

- For whites in poverty:
  - Easier to endorse:
    - Driving while drinking ($\Delta\kappa$ = - 0.095).
    - Driving after drinking ($\Delta\kappa$ = - 0.054).
  - More difficult to endorse alcohol:
    - Caused trouble with job/school ($\Delta\kappa$ = 0.100).
    - Caused trouble with family/friends ($\Delta\kappa$ = 0.058).
    - Led to fights ($\Delta\kappa$ = 0.067)
    - Led to legal problems ($\Delta\kappa$ = 0.064)
- Similar (but not identical) pattern among Blacks and Hispanics.

Cincinnati Children's

Results: MG-MIMIC

- Next examined equivalence in the loadings across Whites, Blacks, and Hispanics.

- Uncovered systematically biased loadings.

  - $\Delta\chi^2 = 30.40$ (14, $n = 25{,}512$, $p < 0.01$)

Results: MG-MIMIC

- For Blacks:
  - Driving while drinking related less strongly to abuse ($\Delta\lambda$ = - 0.852).
- For Hispanics:
  - Driving while drinking related less strongly to abuse ($\Delta\lambda$ = - 0.903).
  - Riding in vehicle while driver drinks related less strongly to abuse ($\Delta\lambda$ = - 0.419).
- For the driving while drinking item, the loading did not differ between Blacks and Hispanics.

Cincinnati Children's

Results: MG-MIMIC

- Next examined equivalence in the thresholds across Whites, Blacks, and Hispanics.

- Model uncovered systematically biased thresholds.

  - $\Delta\chi^2 = 88.87$ (13, n = 25,512, $p < 0.01$)

Results: MG-MIMIC

- For Blacks:
  - Easier to endorse drinking while driving (Δv = - 0.921).
  - More difficult to endorse driving after drinking (Δv = 0. 593).
  - More difficult to endorse entering harmful situations after drinking (Δv = 0.68).
- For Hispanics:
  - Easier to endorse drinking while driving (Δv = - 0.96).
  - More difficult to endorse driving after drinking too much (Δv = 0.457).
  - More difficult to endorse entering harmful situations after drinking (Δv = 0.456).

Cincinnati Children's

Results: MG-MIMIC

- Compared estimates ignoring systematic measurement error to adjusted estimates from final MG-MIMIC model.
  - Adjusted estimates mitigate measurement error.

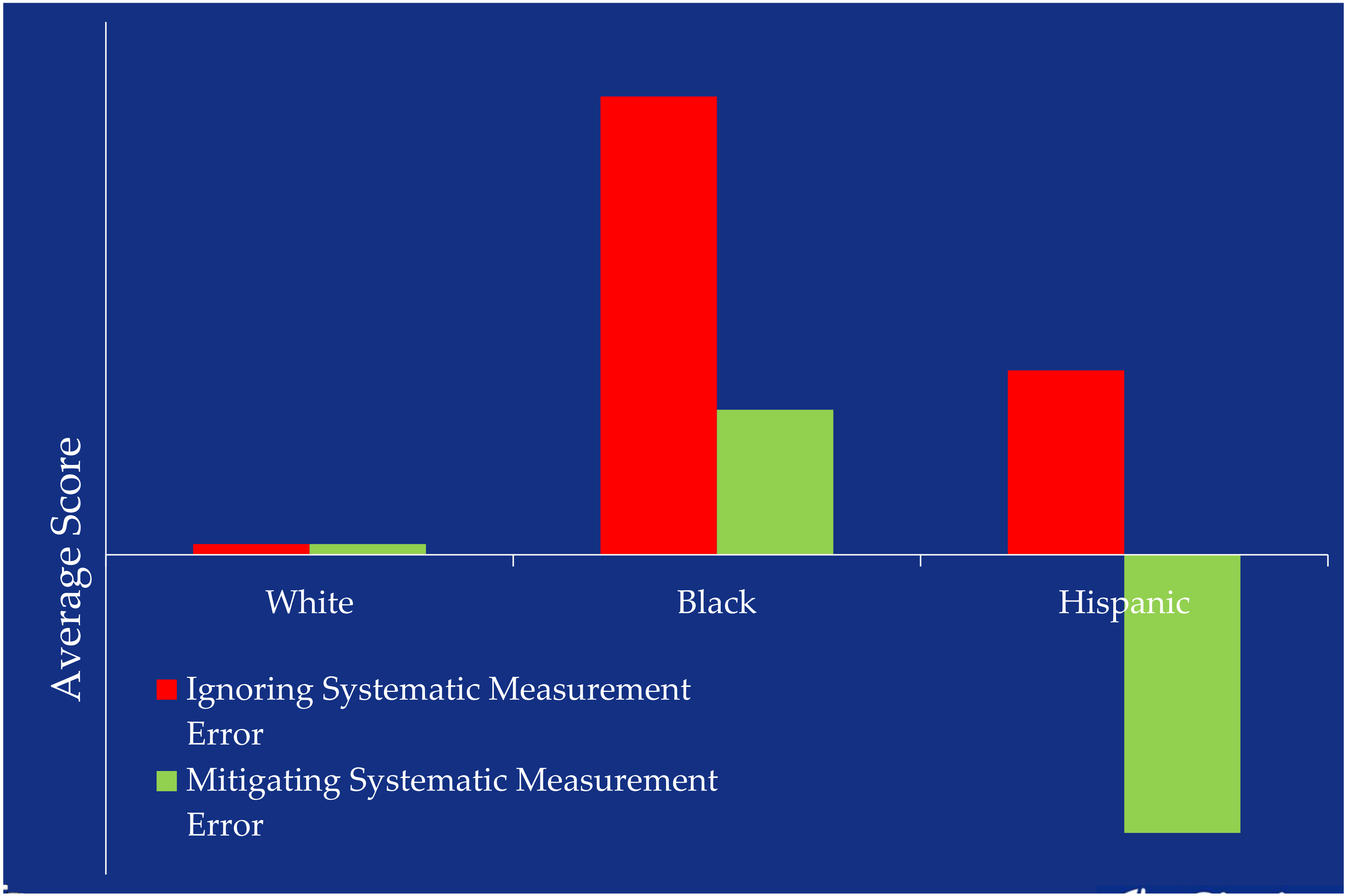
- Addresses whether measurement bias influences conclusions.

Results: MG-MIMIC

- Unadjusted means:
  - Whites: 0 (Reference group)
  - Blacks: 0.43 z = 7.82, $p < 0.01$: More abuse.
  - Hispanics : 0.173 = *2.11*, $p < 0.05$: More abuse.

- Adjusted means:
  - Whites: 0 (Reference group)
  - Blacks: 0.136 z = 1.855, *ns*.: No difference.
  - Hispanics : -2.261 = *-2.20*, $p < 0.05$: Less abuse.

Results: MG-MIMIC



Discussion

- MG-MIMIC showed that systematic measurement error significantly and substantially affects alcohol abuse affects estimates in the diverse population.
- Without mitigating systematic error, efforts to identify and understand disparities and inequities across these populations result in flawed conclusions.

Discussion

- Ignoring systematic measurement error:
  - Appears both Blacks and Hispanics engage in more alcohol abuse behavior relative to Whites.
    - Based on unadjusted estimates.

- After mitigating systematic measurement error, analyses show that:
  - Blacks do not differ from Whites.
  - Hispanics engage in less abuse behavior than Whites.

Cincinnati Children's

Limitations

- Self report data.
  - Responses may not reflect children's actual health.

Limitations

- MG-MIMIC can only detect biased thresholds for background variables.
  - Leaves loadings unaddressed.
  - Problem for background variables *only*.
  - Missing data approach possibilities.
- If bias permeates the *entire* question set, analyses cannot detect this.
  - No statistical approach can.
    - (Millsap, 2006).

Conclusion

- Investigators too often treat race and ethnicity as explanatory variables.
- Ethnicity acts as a proxy for other variables that systematically vary across people of different ethnic backgrounds.

Conclusion

- We should seek to uncover the variables for which ethnicity serves as a proxy.

- We should advance our statistical models to incorporate the multiple influences on health outcomes.

Conclusion

- We should seek to uncover the variables for which ethnicity serves as a proxy.

- We should advance our statistical models to incorporate the multiple influences on health outcomes.

Conclusion

- Remember, before making cross-group comparisons, *must* consider measurement equivalence across groups.

- Do *observed* group differences reflect *true* differences?

- Or, do group differences result from systematic measurement error?

- And, do *observed* similarities reflect *true* similarities?

References

- For detailed presentation:
- Carle, A. C. (2010). Mitigating systematic measurement error in comparative effectiveness research in heterogeneous populations. *Medical Care*.
- References available upon request.


TUT TUT...
'NUF SAID!
SIDDOWN...
RELAX...
©R.CRUMB

Systematic Measurement Error's Influence on Estimating and Understanding Health Disparities

Adam C. Carle, M.A., Ph.D.
adam.carle@cchmc.org
James M. Anderson Center for Health Systems Excellence
Cincinnati Children's Hospital Medical Center
University of Cincinnati School of Medicine
University of Cincinnati College of Arts and Sciences