THE GOOD, THE BAD, & THE UGLY: WHAT WE KNOW TODAY ABOUT LCA WITH DISTAL OUTCOMES

Bethany C. Bray, Ph.D.
bcbray@psu.edu

WHAT ARE WE HERE TO TALK ABOUT TODAY?

Behavioral scientists increasingly are using latent class analysis (LCA) to identify subgroups of individuals based on unique patterns of…

- Behavior
- Risk exposure
- Mental health symptoms
- Other characteristics

LCA is a powerful and intuitive tool for studying heterogeneity in these characteristics

WHAT ARE WE HERE TO TALK ABOUT TODAY?

But, new methods are needed to address the next generation of complex questions

How is subgroup membership embedded in developmental pathways?

For example, how subgroup membership is linked to later outcomes

- Do patterns of early risk exposure during childhood predict later binge drinking during adolescence?
- Do patterns of depression symptoms during adolescence predict later academic achievement?

WHAT ARE WE HERE TO TALK ABOUT TODAY?

LCA with a distal outcome poses interesting methodological challenges

Literature has included a rapidly increasing number of publications proposing competing approaches to address these challenges

Summarize three state-of-the-art approaches to LCA with distal outcomes

- Focus on the simplest case of a latent class predictor and an observed distal outcome

OUTLINE OF TODAY'S TALK

Why latent class analysis (LCA)?

- A brief introduction to LCA

Why are distal outcomes so troublesome?

- Traditional approaches to LCA with distal outcomes
- Contemporary approaches to LCA with distal outcomes

How are researchers supposed to choose?

- The good, the bad, and the ugly
- Take-home messages

OUTLINE OF TODAY'S TALK

Why latent class analysis (LCA)?

- A brief introduction to LCA

Why are distal outcomes so troublesome?

- Traditional approaches to LCA with distal outcomes
- Contemporary approaches to LCA with distal outcomes

How are researchers supposed to choose?

- The good, the bad, and the ugly
- Take-home messages

WHY LCA?

Statistical tool that behavioral scientists are turning to with increasing frequency

Can be used to explain population heterogeneity by identifying underlying subgroups of individuals

Subgroups (classes) are comprised of individuals who are similar in their responses to a set of observed variables

Class membership is inferred from responses to the observed variables

OUTLINE OF TODAY'S TALK

Why latent class analysis (LCA)?

- **A brief introduction to LCA**

Why are distal outcomes so troublesome?

- Traditional approaches to LCA with distal outcomes
- Contemporary approaches to LCA with distal outcomes

How are researchers supposed to choose?

- The good, the bad, and the ugly
- Take-home messages

A (VERY) BRIEF INTRODUCTION

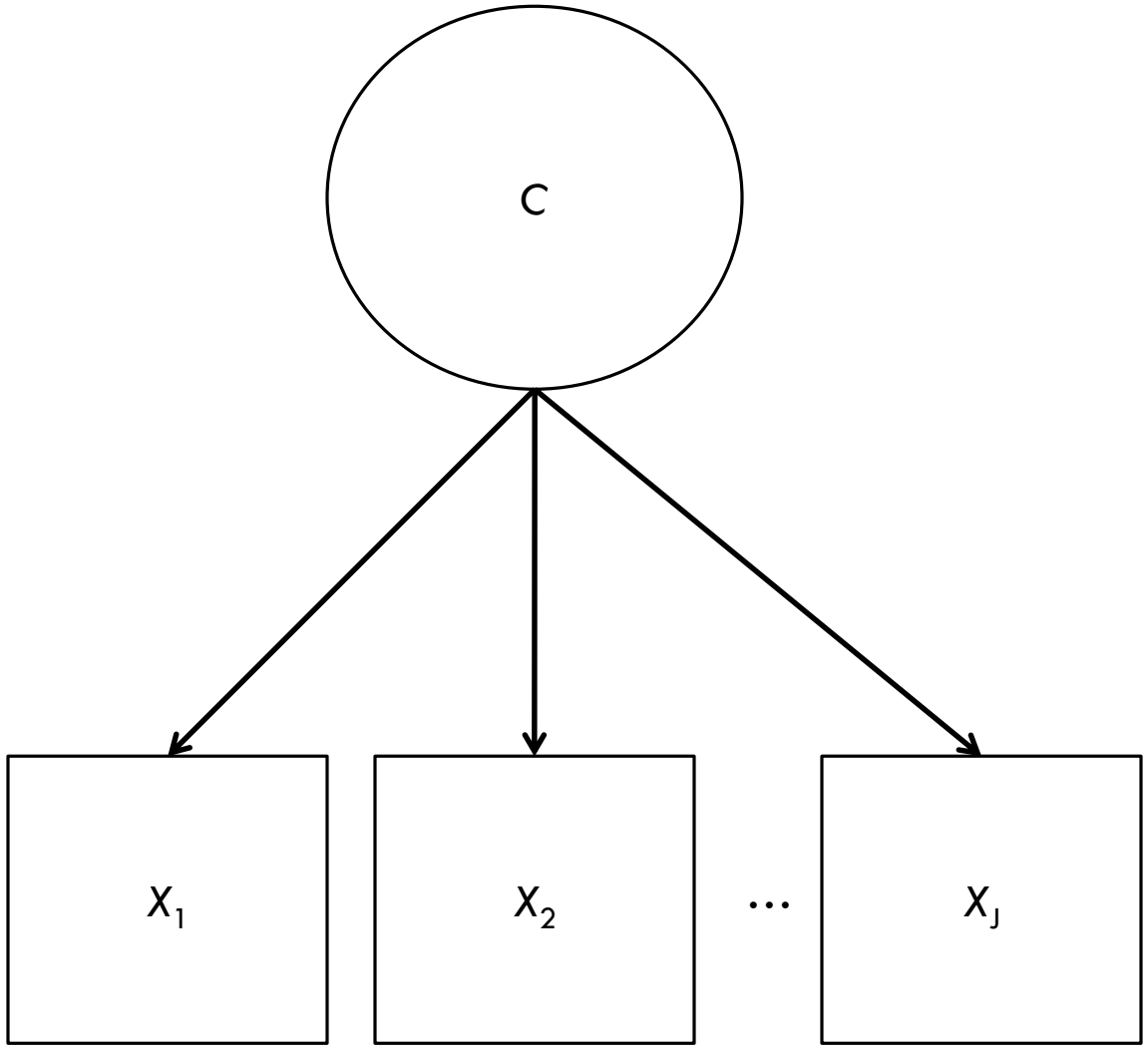
LCA posits mutually exclusive and exhaustive underlying set of latent classes

- Classes and class membership inferred from multiple categorical observed variables

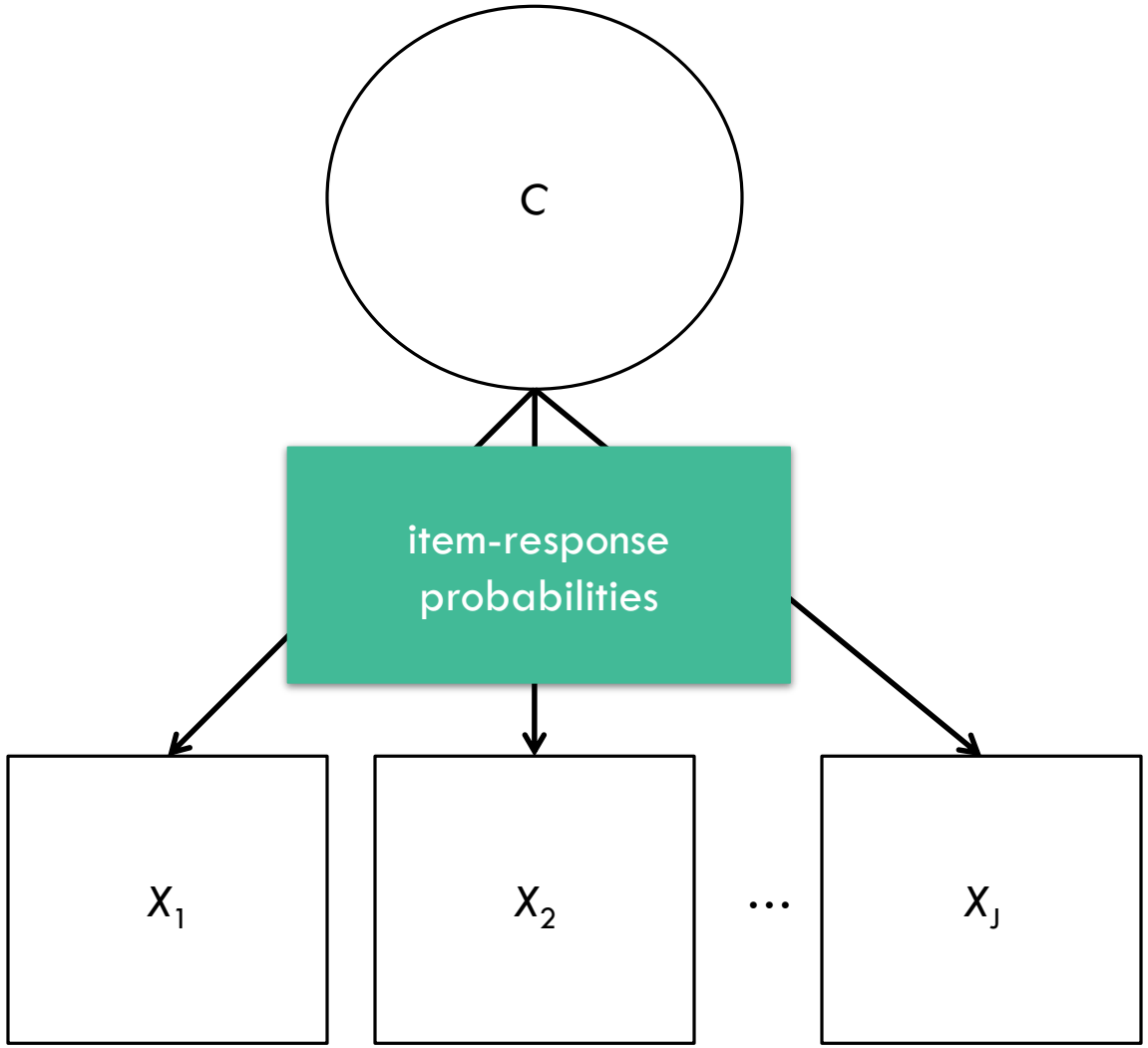
In traditional model, interested in two sets of parameters…

- Rhos: item-response probabilities
- Gammas: latent class membership probabilities

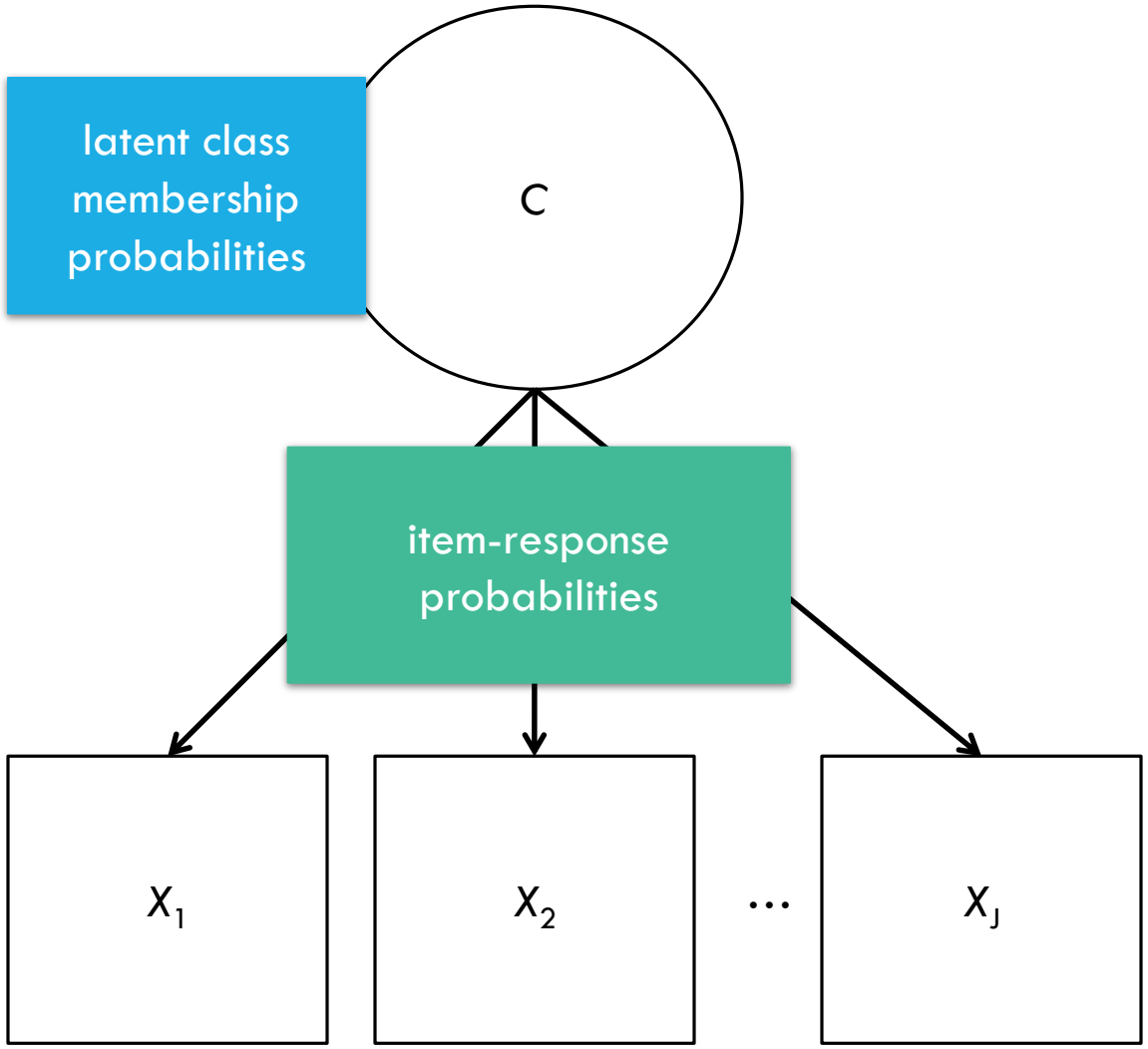
A (VERY) BRIEF INTRODUCTION



A (VERY) BRIEF INTRODUCTION



A (VERY) BRIEF INTRODUCTION



A (VERY) BRIEF INTRODUCTION

Often interested in understanding what predicts latent class membership

- What are the important predictors of patterns of early risk exposure during childhood?
- What are the important predictors of patterns of depression symptoms during adolescence?
- More concrete: do adolescents' friendship goals (i.e., a risk factor) predict substance use patterns (i.e., a latent class variable)?

A (VERY) BRIEF INTRODUCTION

Mathematical model for predicting class membership from a covariate is well-understood

Estimating the association between a latent class predictor and distal outcome presents a more difficult methodological problem

Solving this problem is a “hot topic” in the methodological literature right now

Three competing state-of-the-art approaches to LCA with distal outcomes

OUTLINE OF TODAY'S TALK

Why latent class analysis (LCA)?

- A brief introduction to LCA

Why are distal outcomes so troublesome?

- Traditional approaches to LCA with distal outcomes
- Contemporary approaches to LCA with distal outcomes

How are researchers supposed to choose?

- The good, the bad, and the ugly
- Take-home messages

PENNSTATE
The Methodology Center
advancing methods, improving health

WHY LCA WITH DISTAL OUTCOMES?

What if we are interested in…

- Predicting later binge drinking from early risk exposure
- Predicting later academic achievement from depression subtypes

One modeling option is to use…

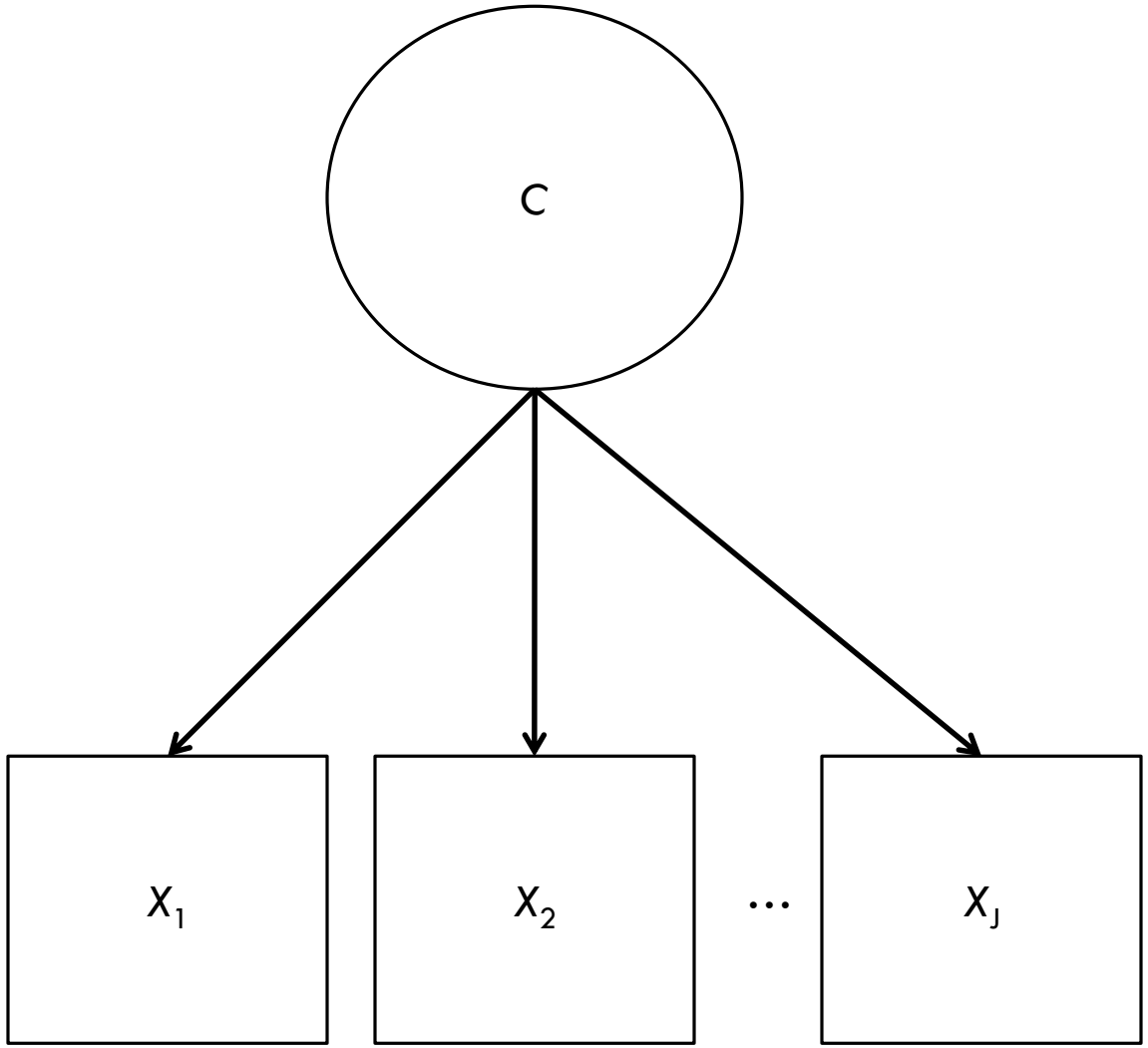
- Latent class analysis (LCA) with a distal outcome
  - Latent class variable is risk exposure
  - Distal outcome is observed binge drinking

WHY LCA WITH DISTAL OUTCOMES?

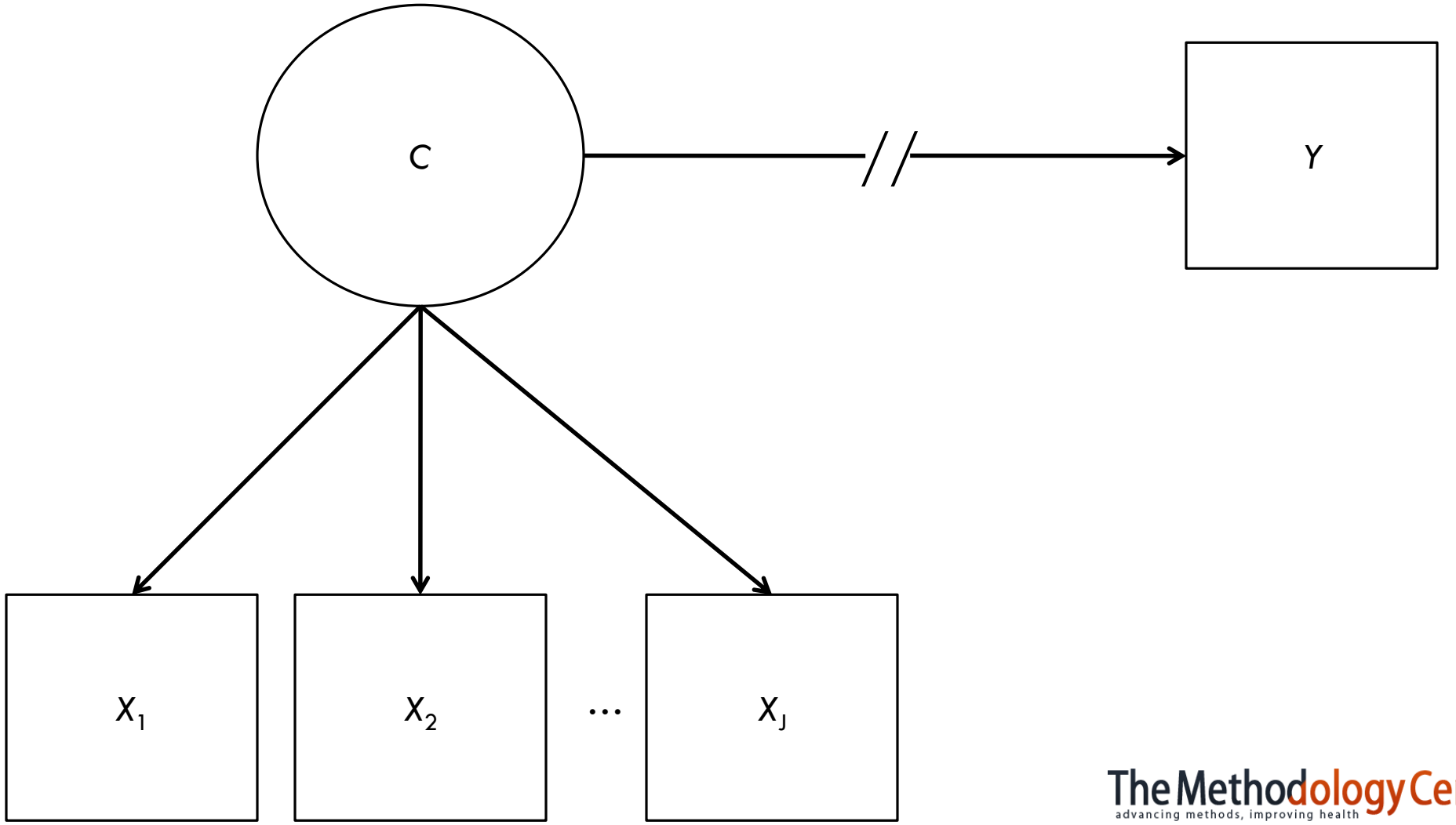
When using latent class membership to predict a distal outcome, interested in effect of *C* on *Y*

Let's think about this graphically…

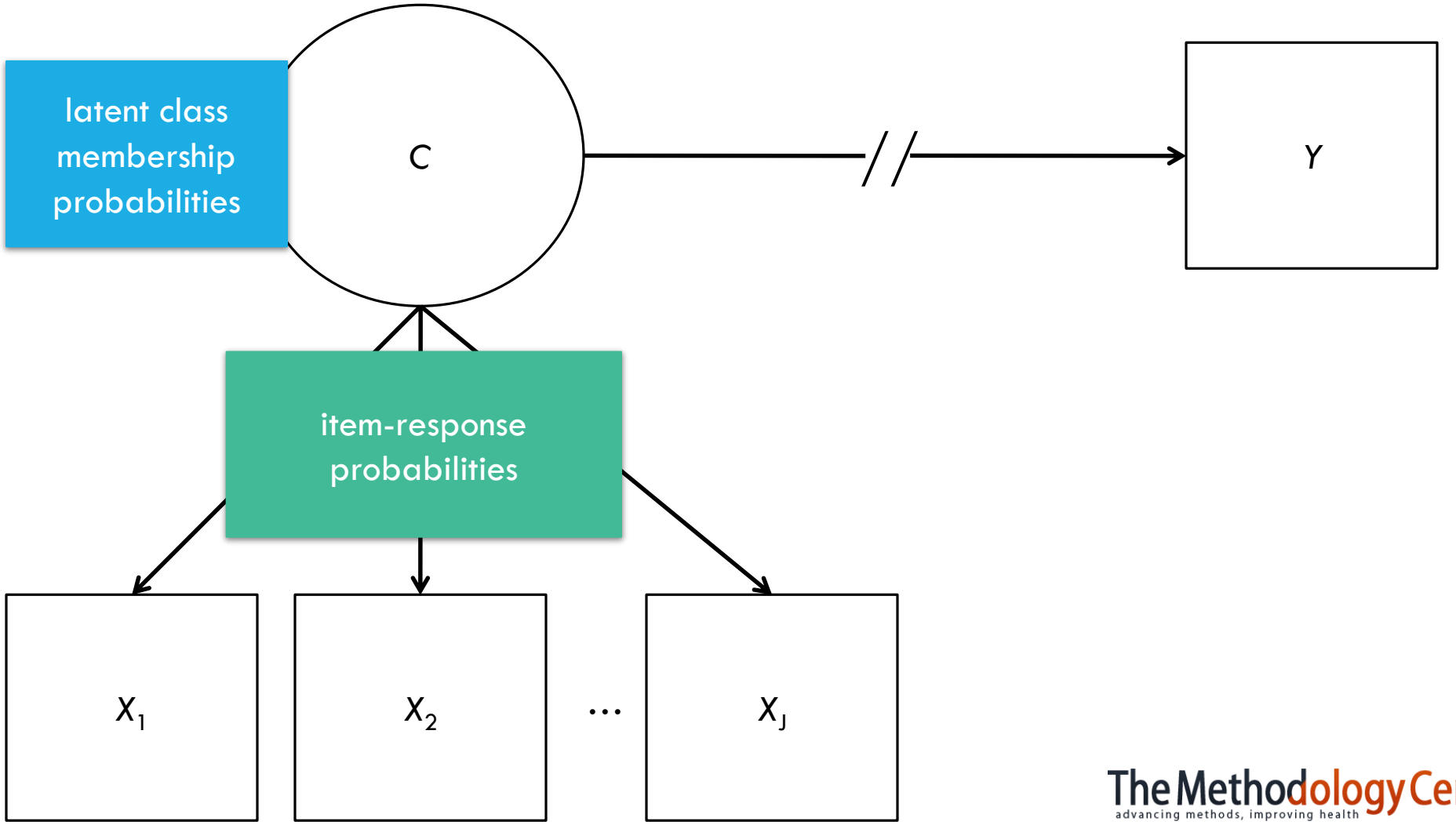
WHY LCA WITH DISTAL OUTCOMES?



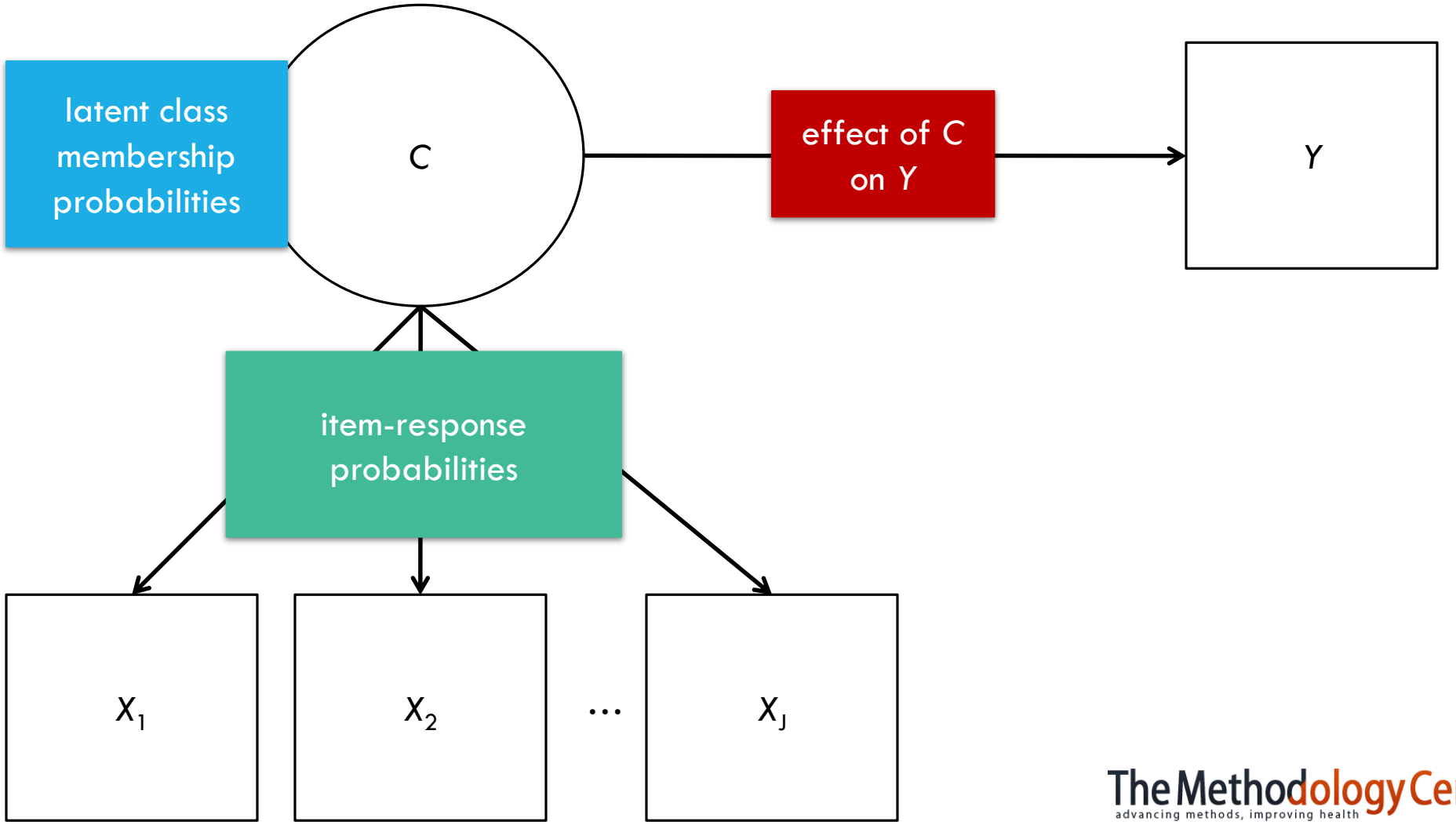
WHY LCA WITH DISTAL OUTCOMES?



WHY LCA WITH DISTAL OUTCOMES?



WHY LCA WITH DISTAL OUTCOMES?



WHY LCA WITH DISTAL OUTCOMES?

For example, what is the effect of risk exposure latent class membership on binge drinking?

- Classes of individuals…
  - Low Risk
  - Peer Risk
  - Economic Risk
  - Household & Peer Risk
  - Multi-Risk

WHY LCA WITH DISTAL OUTCOMES?

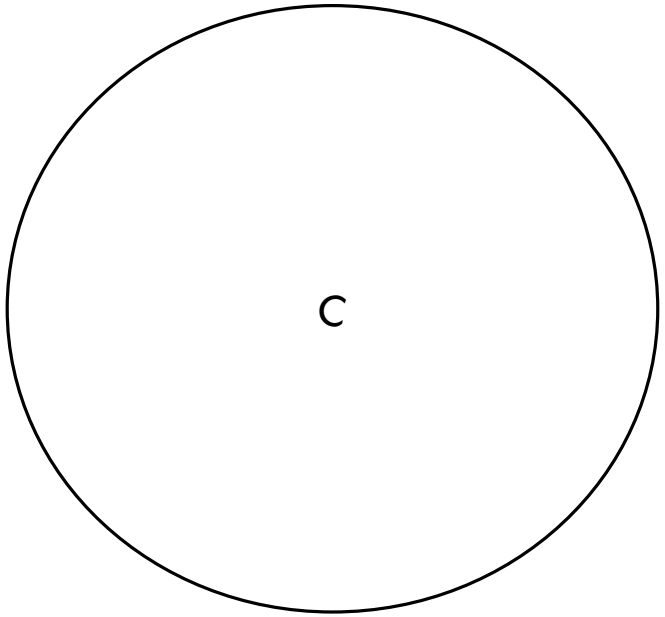
For example, what is the effect of risk exposure latent class membership on binge drinking?

- Classes of LOW RISK and HIGH RISK individuals
- Does prevalence of BINGE DRINKING differ between low risk and high risk individuals

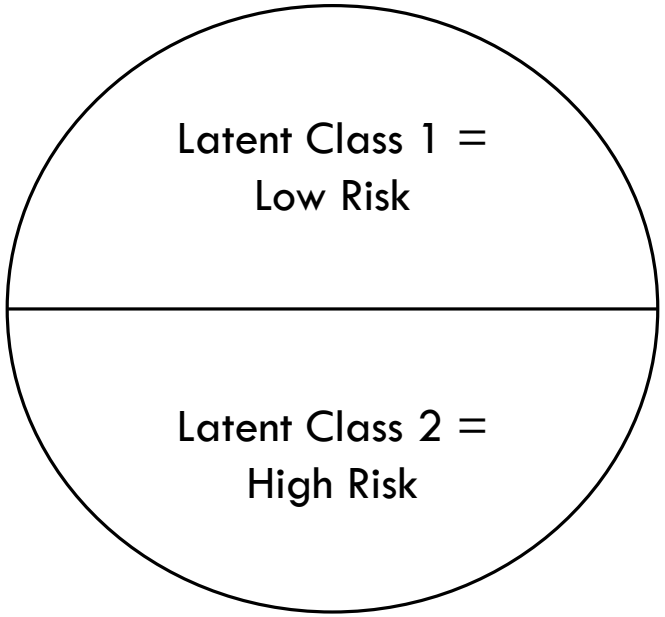
To address this question, we need to know the conditional distribution of *Y* given C

- That is, the probabilities of binge drinking for both low risk and high risk individuals

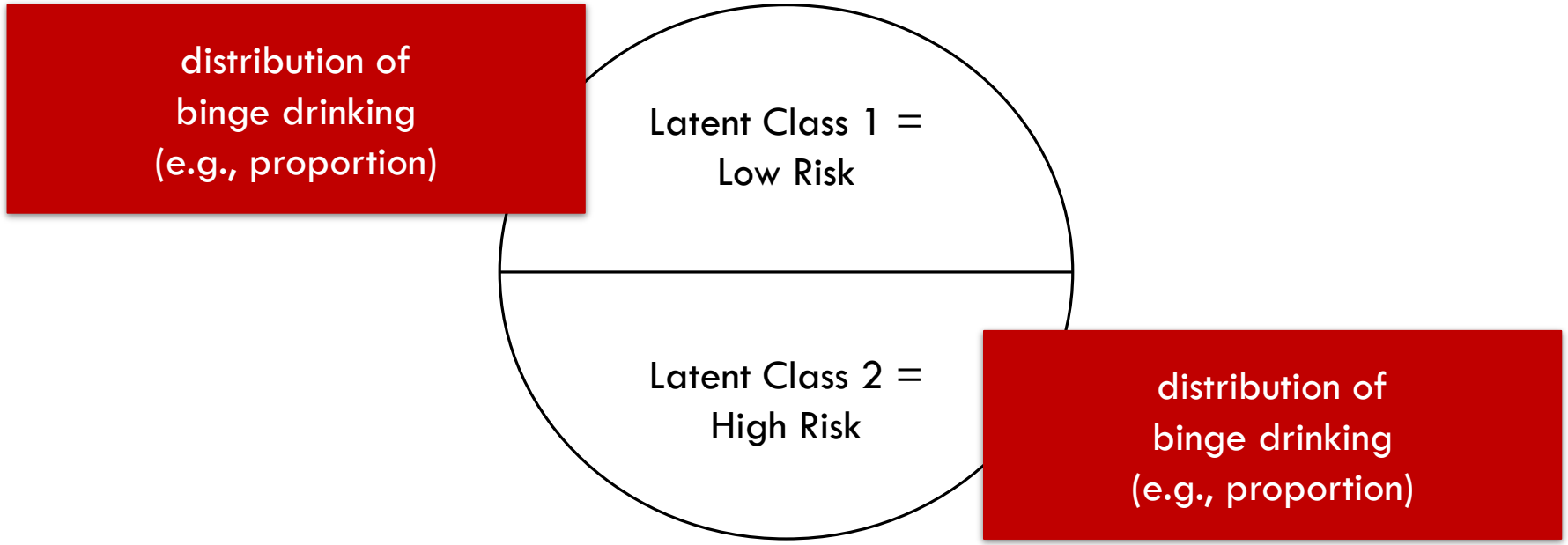
WHY LCA WITH DISTAL OUTCOMES?



WHY LCA WITH DISTAL OUTCOMES?



WHY LCA WITH DISTAL OUTCOMES?

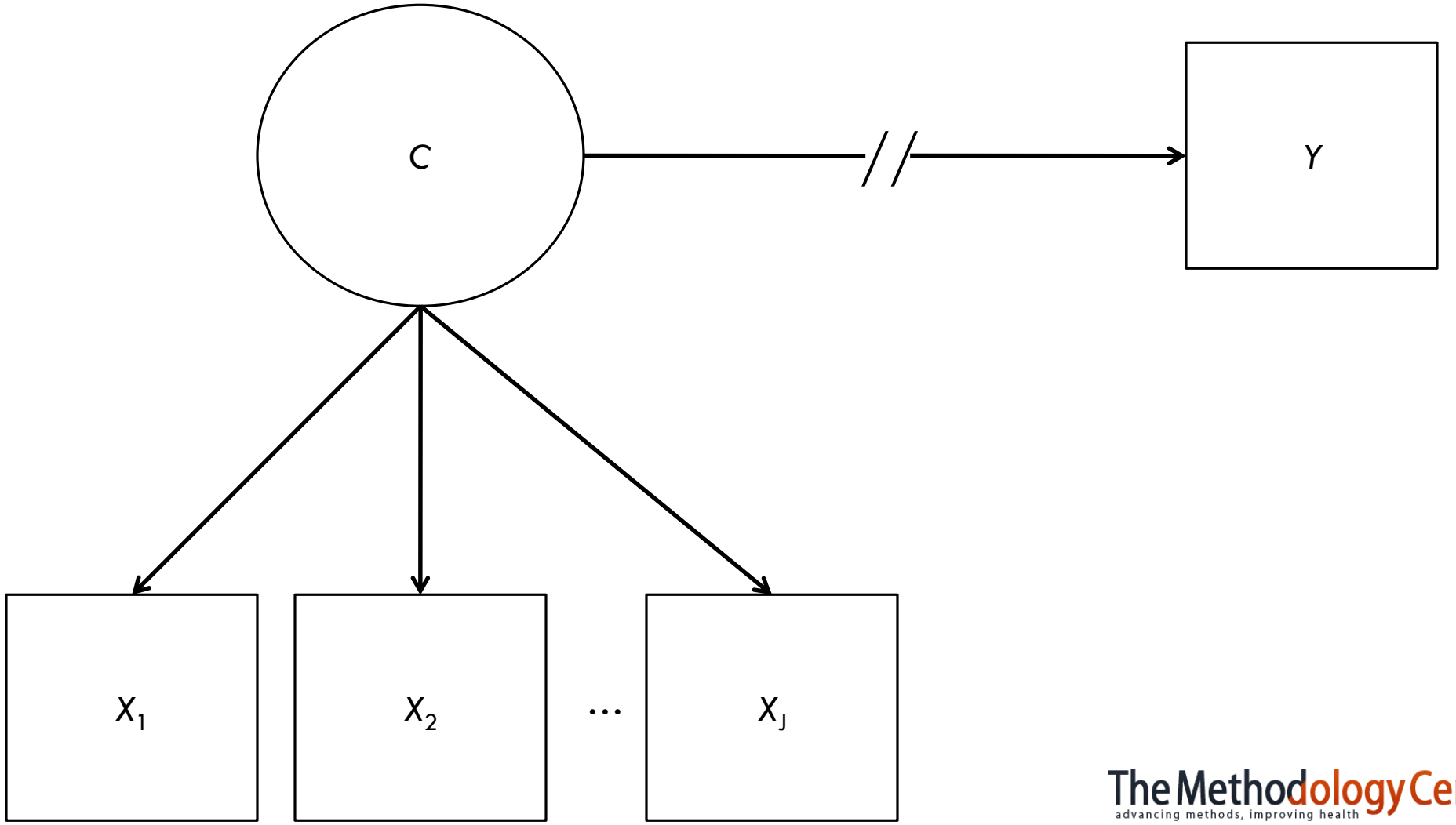


WHY ARE DISTAL OUTCOMES SO TROUBLESOME?

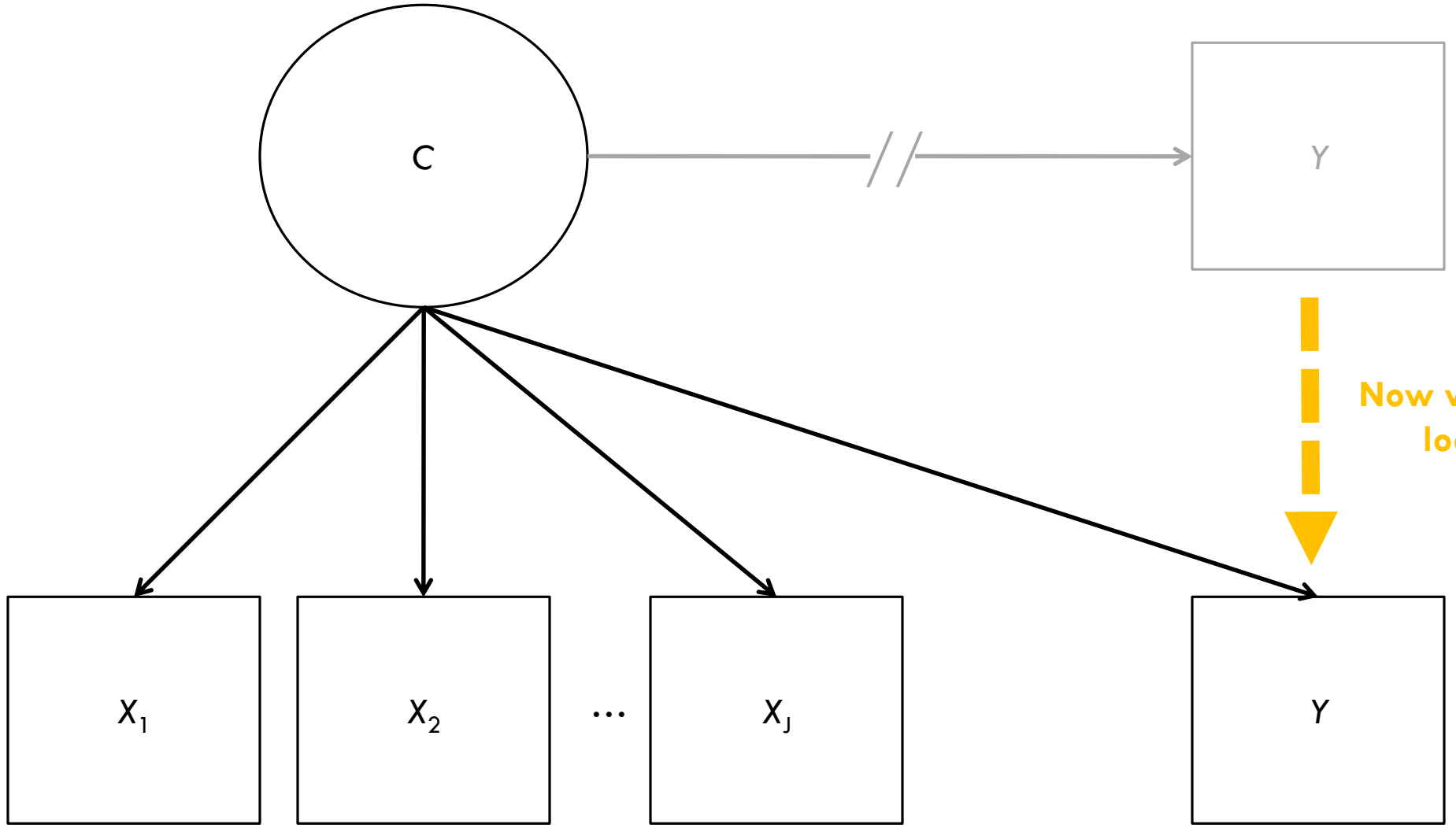
LCA with a distal outcome poses interesting methodological challenges

But, why?

WHY ARE DISTAL OUTCOMES SO TROUBLESOME?



WHY ARE DISTAL OUTCOMES SO TROUBLESOME?



WHY ARE DISTAL OUTCOMES SO TROUBLESOME?

How do we get $Y|C$?

WHAT IS THE SOLUTION?

Historically, classify-analyze strategies have been used to solve this problem

Individuals are assigned to classes using some rule based on posterior probabilities

Then an outcome analysis is performed treating class membership as known

- For example, regressing the outcome on a set of dummy coded predictors for class assignment

This approach, however, is known to cause substantial attenuation in effect estimates

WHAT IS THE SOLUTION?

Instead, there are three general categories of state-of-the-art approaches to LCA with distal outcomes

- Each has been shown to work well under certain conditions in simulation studies
- How do scientists make informed decisions about which to choose?

OUTLINE OF TODAY'S TALK

Why latent class analysis (LCA)?

- A brief introduction to LCA

Why are distal outcomes so troublesome?

- **Traditional approaches to LCA with distal outcomes**
- Contemporary approaches to LCA with distal outcomes

How are researchers supposed to choose?

- The good, the bad, and the ugly
- Take-home messages

TRADITIONAL APPROACHES

These approaches are based on a simple idea…

Finding $Y|C$ is difficult because C is unobserved

So, make C observed and then find $Y|C$

These approaches are often called classify-analyze or 3-step approaches

TRADITIONAL APPROACHES

All classify-analyze approaches rely on posterior probabilities

- Each individual has posterior probability of membership for latent class: $P(C = c \mid Y = y)$

Use posterior probabilities as basis for classify-analyze approaches

- Classification step:
  Classify individuals to latent classes based on probabilities

- Analysis step:
  Treat latent class membership as known in analysis model

TRADITIONAL APPROACH #1

Modal or maximum-probability assignment

- Fit and compare competing LCAs to select optimal model
- Calculate posterior probabilities for each individual, for each latent class
- Assign individuals to latent class with highest posterior probability
- Conduct analysis by regressing distal outcome on latent class membership

TRADITIONAL APPROACH #2

Proportional assignment

- Like modal assignment
- But, “partially” assign individuals based on their posterior probability distributions
- Conduct analysis by regressing distal outcome on latent class membership

TRADITIONAL APPROACH #3

Multiple pseudo-class draws

- Select optimal model
- Calculate posterior probabilities
- Assign individuals based on distribution of posterior probabilities
- Conduct analysis
- Repeat steps 3 & 4 multiple (e.g., 20) times
- Combine results using rules from multiple imputation

TRADITIONAL APPROACHES

The regression model gives us $Y|C$

- Again, this conditional distribution is what we care about…
- Does the distribution of BINGE DRINKING differ across latent classes

But, numerous simulation studies have shown that these approaches severely attenuate the estimated relation between C and Y

HOW BAD IS BAD?

Effect Size	Max-Prob Non-Inclusive	Max-Prob Inclusive	Pseudo-class Non-Inclusive	Pseudo-class Inclusive
Large	**-.156**	.041	**-.191**	.001
Medium	**-.083**	.028	**-.103**	.006
Small	**-.031**	.009	**-.039**	.001
No effect	**.000**	.000	**.000**	.000

OUTLINE OF TODAY'S TALK

Why latent class analysis (LCA)?

- A brief introduction to LCA

Why are distal outcomes so troublesome?

- Traditional approaches to LCA with distal outcomes
- **Contemporary approaches to LCA with distal outcomes**

How are researchers supposed to choose?

- The good, the bad, and the ugly
- Take-home messages

CONTEMPORARY APPROACHES

Contemporary approaches to LCA with distal outcomes are grouped into three main types…

Weighting by classification error

Bayes' theorem based approach

Posterior probability improvement

CONTEMPORARY APPROACHES

Each approach has been shown to work well under certain conditions in recently published simulation studies

However, there has been no comprehensive overview summarizing the approaches and their assumptions or integration of "take-home messages" across simulation studies

To further complicate matters, not all approaches are implemented in all LCA software packages and the availability of high-quality standard errors depends on the combination of approach and software package selected

WEIGHTING BY CLASSIFICATION ERROR

Assign individuals to classes based on responses to indicators only

- Assignment typically uses modal or proportional assignment

Retain information about the classification error rate

Treat assignments as known in a subsequent analysis model weighted by the classification error rate

Bakk, Z., & Vermunt, J. K. (in press). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*.

BAYES' THEOREM BASED APPROACH

Fit latent class model and include distal outcome as a covariate

Use Bayes' theorem to reverse the direction of the effect, empirically derive distribution of the distal outcome given class membership

- Distribution based directly on the observed data or using a smoothed or un-smoothed kernel density estimator

Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(1), 1-26.

POSTERIOR PROBABILITY IMPROVEMENT

Assign individuals to classes based on responses to indicators and distal outcome

- Assignment typically uses modal assignment or multiple pseudo-class draws

Then treat assignments as known in subsequent analysis

Bray, B. C., Lanza. S. T., & Tan, X. (2015). Eliminating bias in classify-analyze approaches for latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(1), 1-11.

HOW GOOD IS GOOD?

Effect Size	Max-Prob Non-Inclusive	Max-Prob Inclusive	Pseudo-class Non-Inclusive	Pseudo-class Inclusive
Large	-.156	**.041**	-.191	**.001**
Medium	-.083	**.028**	-.103	**.006**
Small	-.031	**.009**	-.039	**.001**
No effect	.000	**.000**	.000	**.000**

PENNSTATE
The Methodology Center
advancing methods, improving health

WHAT CAN THIS LOOK LIKE IN REAL LIFE?

Approach	Low Risk	Peer Risk	Economic Risk	H.Hold & Peer Risk	Multi-Risk
Max-Prob Non-Inc	.16	.39	.18	**.38**	.44
Pseudo-class Non-Inc	.16	.37	.17	**.39**	.41
Max-Prob Inclusive	.11	.42	.12	**.60**	.36
Pseudo-class Inclusive	.11	.41	.12	**.62**	.36

OUTLINE OF TODAY'S TALK

Why latent class analysis (LCA)?

- A brief introduction to LCA

Why are distal outcomes so troublesome?

- Traditional approaches to LCA with distal outcomes
- Contemporary approaches to LCA with distal outcomes

How are researchers supposed to choose?

- The good, the bad, and the ugly
- Take-home messages

OUTLINE OF TODAY'S TALK

Why latent class analysis (LCA)?

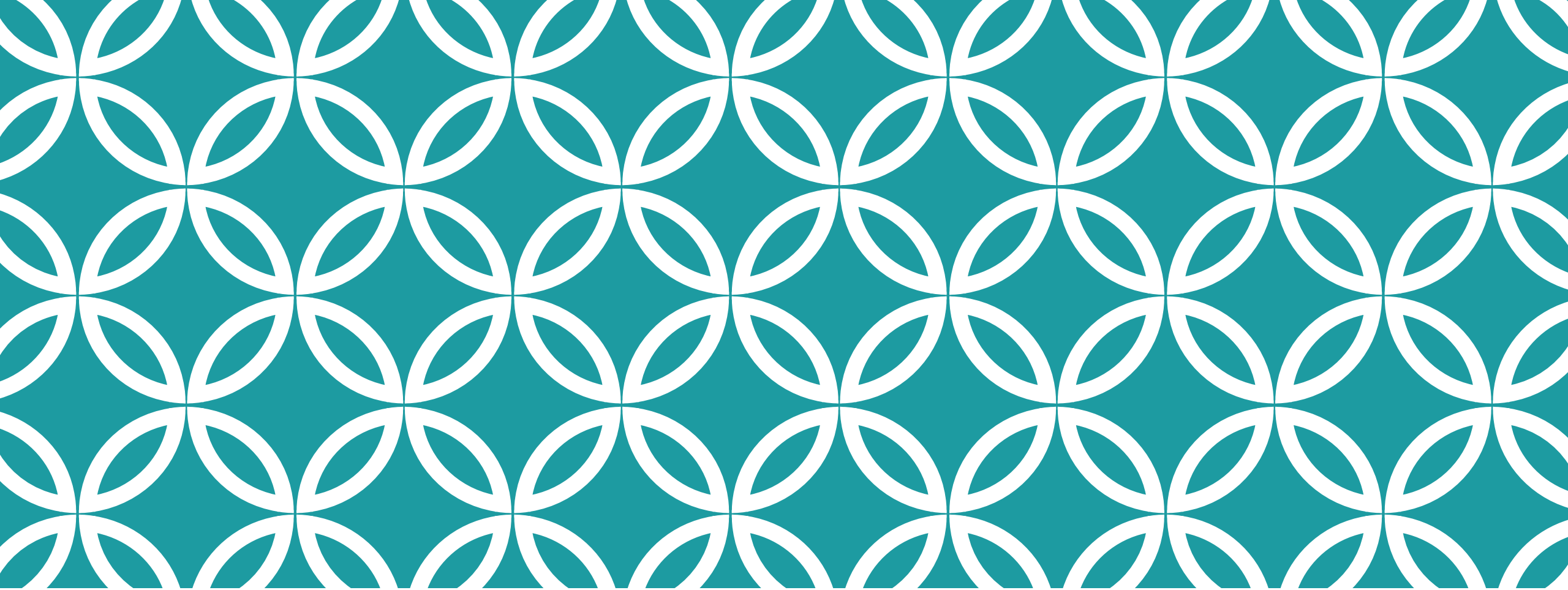
- A brief introduction to LCA

Why are distal outcomes so troublesome?

- Traditional approaches to LCA with distal outcomes
- Contemporary approaches to LCA with distal outcomes

How are researchers supposed to choose?

- **The good, the bad, and the ugly**
- Take-home messages



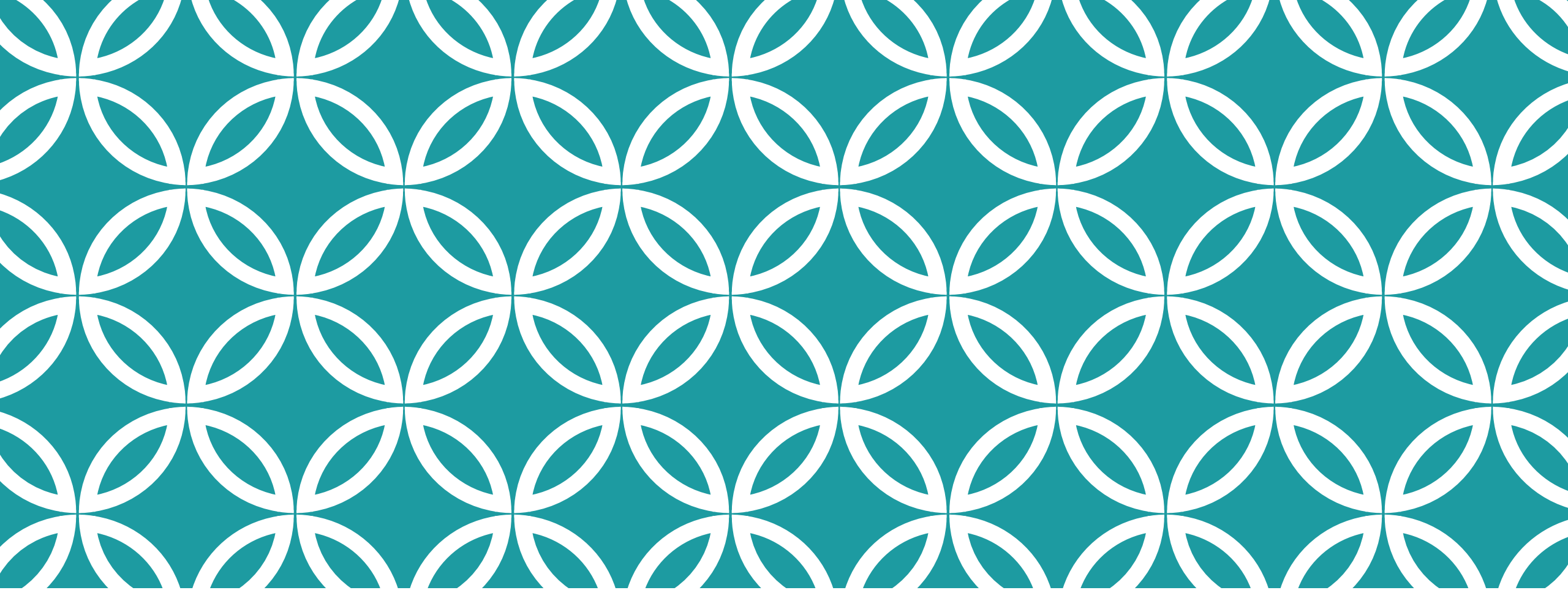
THE GOOD

THE GOOD

All contemporary approaches substantially reduce attenuation and can result in unbiased estimates

Software of some kind is available for all contemporary approaches

- Latent Gold
- Mplus
- SAS



THE BAD

THE BAD

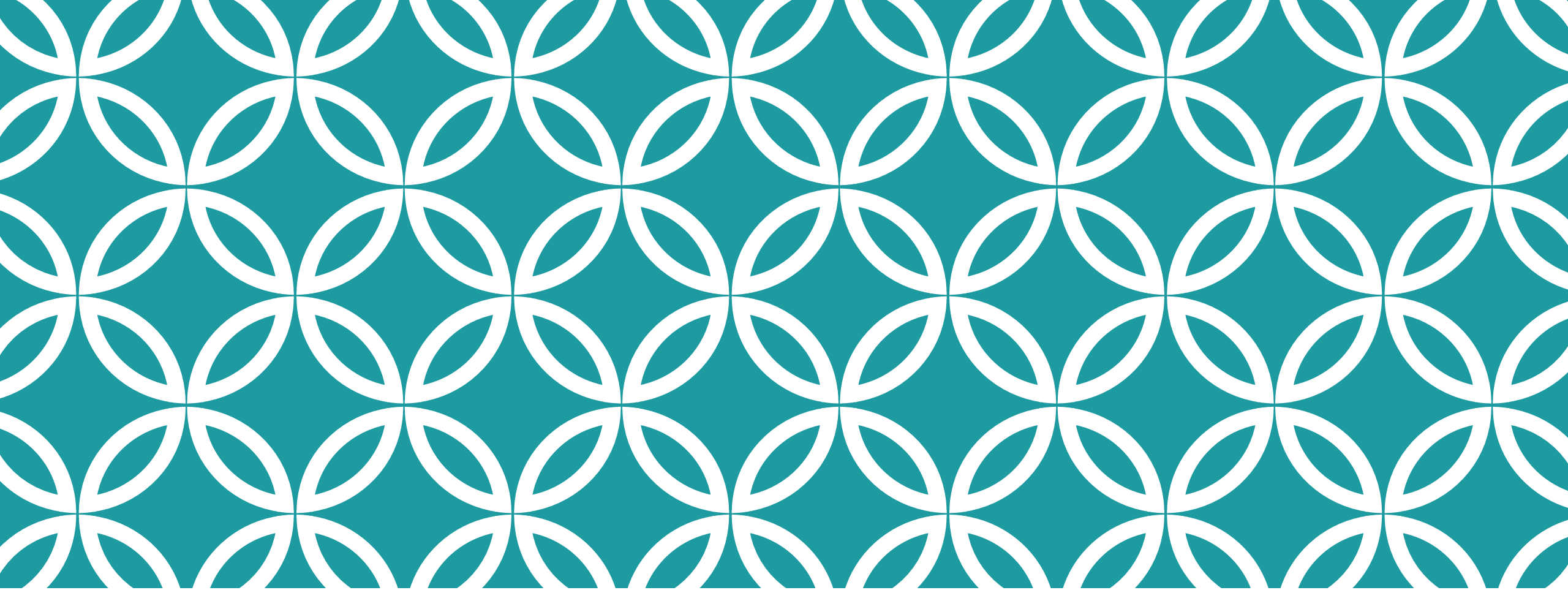
All approaches DO NOT work with all types of outcomes

No single software package accommodates all of the approaches

All approaches seem to be highly sensitive to violations of model assumptions

Standard errors for all approaches need work

- But, bootstrapping seems most promising



THE UGLY

	PROS TO APPROACH						
APPROACH	Reduces bias in estimates	High-quality std. errs. readily available	Can implement in any LCA software	Msrmnt model does not change across analyses	Does not require assigning individuals	Robust to violations of analysis model assumptions	Complexity of analysis model unlimited
Traditional	No	No	✓	✓	No	✓	✓
#1: Weighting	✓	No	No Latent Gold + Mplus	✓	No	✓	No
#2: Bayes	✓	No	No SAS + Mplus	No	✓	No but can be improved	No
#3: Post Probs	✓	No	✓	No	No	No but can be improved	✓

OUTLINE OF TODAY'S TALK

Why latent class analysis (LCA)?

- A brief introduction to LCA

Why are distal outcomes so troublesome?

- Traditional approaches to LCA with distal outcomes
- Contemporary approaches to LCA with distal outcomes

How are researchers supposed to choose?

- The good, the bad, and the ugly
- **Take-home messages**

TAKE-HOME MESSAGES

1. LCA with distal outcomes is a **hot topic** right now

2. There is a universal best approach but, can be difficult to implement

3. Assumptions and standard errors are tricky things

ACKNOWLEDGEMENTS

Collaborators…

John Dziak

Stephanie Lanza

Michael Russell

Xianming Tan

Jie-Ting Zhang

PENNSTATE

The Methodology Center

advancing methods, improving health

ACKNOWLEDGEMENTS

Funding…

The project described was supported by Award Number **P50 DA010075** from the National Institute on Drug Abuse. Content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse or the National Institutes of Health.

Data…

The project described used data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

POST-DOC OPPORTUNITIES

The Methodology Center
advancing methods, improving health

INSTITUTE FOR SOCIAL RESEARCH
UNIVERSITY OF MICHIGAN

Prevention and Methodology Training (PAMT)

- Integrating prevention science and innovative methodology
- NIDA-funded T32

The Methodology Center

Department of Biobehavioral Health, Penn State

- Working with Stephanie Lanza

Institute of Social Research, University of Michigan

- Working with Megan Patrick

PENNSTATE

The Methodology Center
advancing methods, improving health

THANK YOU!!

Bethany C. Bray, Ph.D.

- Research Assistant Professor
- Outreach Director, The Methodology Center
- Training Director, Prevention and Methodology Training (PAMT) Program

bcbray@psu.edu

bethanycbray.wordpress.com

methodology.psu.edu/people/bbray

The Methodology Center
204 East Calder Way, Suite 400
State College, PA 16801

mchelpdesk@psu.edu

methodology.psu.edu

PENNSTATE
The Methodology Center
advancing methods, improving health