Predicting Intersectional Inequalities Using Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA)

Professor George Leckie
Centre for Multilevel Modelling
School of Education
University of Bristol
United Kingdom

g.leckie@bristol.ac.uk

Two questions for the audience!

1. **Do you think MAIHDA is a valid application of multilevel modelling?**

   – From a theoretical perspective things look a little peculiar!

   – But, from a pragmatic perspective MAIHDA appears useful.

2. **What is the best way to demonstrate the greater predictive accuracy of multilevel model predicted intersection means over simple means?**

   – I have plotted analytic expressions.

   – But, I could have conducted a simulation study.

Intersectional inequalities

- Traditional studies of inequalities map mean outcomes across the categories of one sociodemographic characteristic (e.g., gender, ethnicity, SES) at a time

– **Male**	OR	**White**	OR	**Low-SES**
– **Female**		**Black**		**High-SES**

- Intersectional studies of inequalities map mean outcomes across combinations of categories of multiple sociodemographic characteristics.
  - **Male, White, Low-SES**
  - **Male, White, Mid-SES**
  - **...**
  - **Female, Black, High-SES**

- Motivated by intersectionality theory:
  - An interest lies in multiple disadvantage and whether it is non-additive.

Simple approach

- The simple approach is to:
  - Calculate the **simple mean outcome** for each intersection.
  - If we also want to study additivity, we could estimate these means via a **saturated linear regression** on the multiple characteristics and all two-way and higher-order interactions.
- Argued limitations include:
  - **Erratic estimates** due to small sample sizes for rarer combinations.
  - **Overfitting** due to small sample sizes for rarer combinations.
  - **Multiple comparisons** problem due to many intersections.

MAIHDA approach

- The MAIHDA approach assumes **individuals are nested with intersections** and then fits two multilevel models:
  - Model 1: Empty model.
  - Model 2: Includes the sociodemographic characteristics as main effects.
- The intersection means are then **predicted post-estimation**:
  - Empirical Bayes prediction if estimation by frequentist methods (REML).
  - Posterior means if estimation is by Bayesian methods (MCMC).
- These predicted means are argued to be more accurate than simple means due to **shrinkage** (partial pooling).

Social Science & Medicine 203 (2018) 64–73

Contents lists available at ScienceDirect

ELSEVIER

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

A multilevel approach to modeling health inequalities at the intersection of multiple social identities

Clare R. Evans[a,*], David R. Williams[b], Jukka-Pekka Onnela[c], S.V. Subramanian[b]

[a] Department of Sociology, University of Oregon, Eugene, OR, United States
[b] Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, United States
[c] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, United States

ARTICLE INFO

Keywords:
Health inequalities
Intersectionality
Eco-epidemiology
Multilevel modeling

ABSTRACT

Rationale: Examining interactions between numerous interlocking social identities and the systems of oppression and privilege that shape them is central to health inequalities research. Multilevel models are an alternative and novel approach to examining health inequalities at the intersection of multiple social identities. This approach draws attention to the heterogeneity within and between intersectional social strata by partitioning the total variance across two levels.

Method: Utilizing a familiar empirical example from social epidemiology—body mass index among U.S. adults ($N = 32{,}788$)—we compare the application of multilevel models to the conventional fixed effects approach to studying high-dimension interactions. Researchers are often confronted with the need to explore numerous interactions of identities and social processes. We explore the interactions of five dimensions of social identity and position—gender, race/ethnicity, income, education, and age—for a total of 384 unique intersectional social strata.

Results: We find that the multilevel approach provides advantages over conventional models, including scalability for higher dimensions, adjustment for sample size of social strata, model parsimony, and ease of interpretation.

Conclusion: Considerable variation is attributable to the within-strata level, indicating the low discriminatory accuracy of these intersectional identities and the high within-strata heterogeneity of risk that remains unexplained. Multilevel modeling is an innovative and valuable tool for evaluating the intersectionality of health inequalities.

Educational Psychology Review (2023) 35:31
https://doi.org/10.1007/s10648-023-09733-5

REFLECTION ON THE FIELD

Educational Inequalities at the Intersection of Multiple Social Categories: An Introduction and Systematic Review of the Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA) Approach

Lena Keller[1] · Oliver Lüdtke[2,3] · Franzis Preckel[4] · Martin Brunner[1]

Accepted: 10 November 2022 / Published online: 8 March 2023

Abstract
Intersectional approaches have become increasingly important for explaining educational inequalities because they help to improve our understanding of how individual experiences are shaped by simultaneous membership in multiple social categories that are associated with interconnected systems of power, privilege, and oppression. For years, there has been a call in psychological and educational research for quantitative approaches that can account for the intersection of multiple social categories. The present paper introduces the Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA) approach, a novel intersectional approach

Key information

- **Data:** PISA 2018 German subsample
- **Outcome:** Student reading achievement at age 15
- **Level 2:** 40 intersections formed from 4 sociodemog. chars. $(= 2 \times 2 \times 2 \times 5)$:
  - **Gender**: 2 categories (Male, Female)
  - **Immigrant**: 2 categories (Native, Immigrant)
  - **Education**: 2 categories (High-school, University)
  - **Occupation**: 5 categories (Low, Low-Middle, Middle, Middle-High, High)
- **Level 1:** 5,451 students

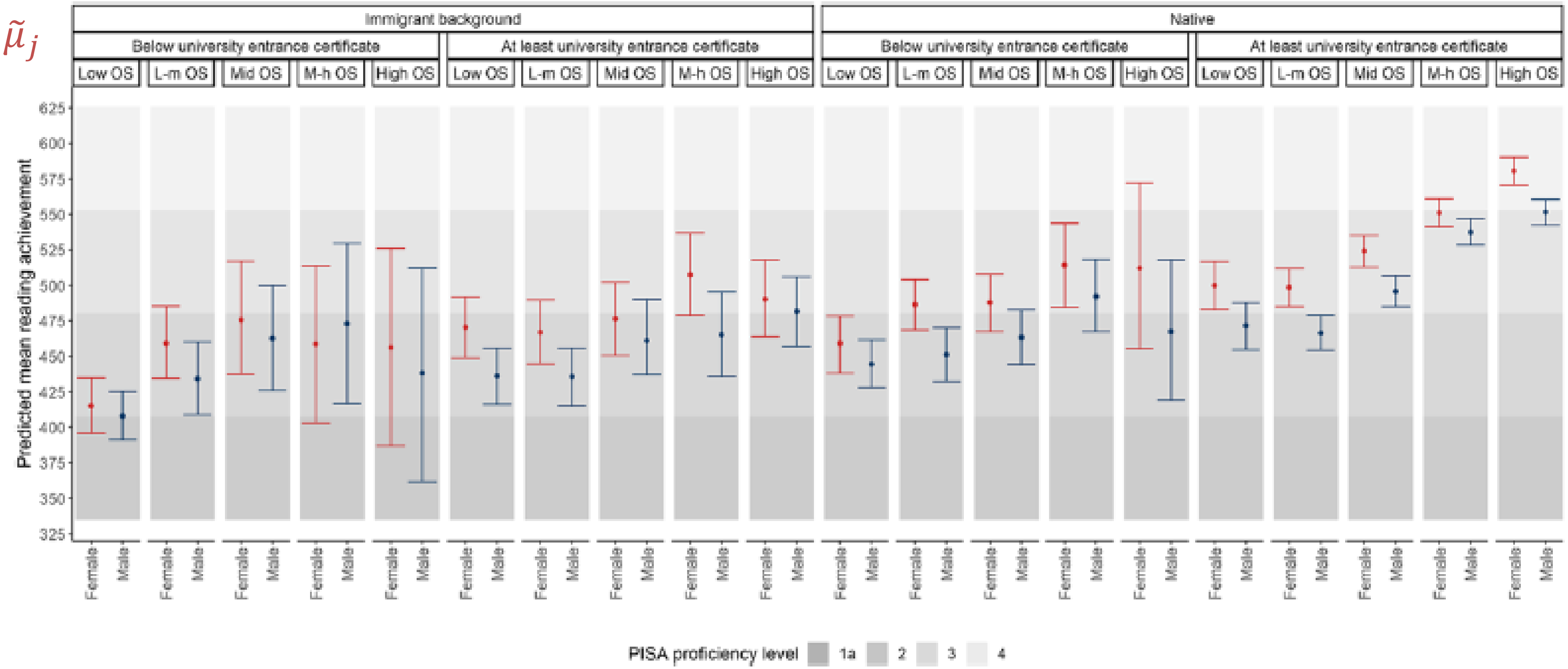
Table 3 Parameter estimates for the multilevel models of reading achievement in 15-year-old students

	Simple intersectional model estimates [95% CI]	Intersectional interaction model estimates [95% CI]
Fixed effects		
Intercept	478.26 [464.27, 492.26]	468.40 [452.81, 483.34]
Gender		
Female (reference)		–
Male		–25.62 [–36.46, –14.91]
Immigrant background		
Native (reference)		–
Immigrant background		–42.46 [–53.71, –30.90]
Highest parental education		
Below university entrance certificate (reference)		–
At least university entrance certificate		29.79 [18.23, 41.99]
Highest parental occupational status (HISEI)		
Low occupational status (reference)		–
Low to middle occupational status		10.06 [–5.85, 26.77]
Middle occupational status		27.63 [11.43, 44.58]
Middle to high occupational status		53.09 [35.19, 70.36]
High occupational status		62.59 [40.50, 81.23]
Measures of variance		
Between-stratum variance	1698.264	144.425
Within-stratum variance	9011.705	9021.200
VPC	15.86%	1.64%
PCV		91.15%

Note. 95% CI = 95% credible intervals; VPC = variance partition coefficient; PCV = proportional change in the between-strata variance

- **MAIHDA Model 1**
- An empty two-level regression
- $\widehat{\mathrm{VPC}} = 0.16$
- 16% of the variation in reading scores is *between* intersections
- 84% of the variation in reading scores is *within* intersections

$\tilde{\mu}_j$



- The predicted intersection means vary greatly (range = 150 points or 1.5 SD).
- The **differential patterns suggests the presence of interactions**:
  - Gender gap varies by combinations of other sociodemographic chars.
  - Occupation gradient varies by combinations of other sociodemog. chars.

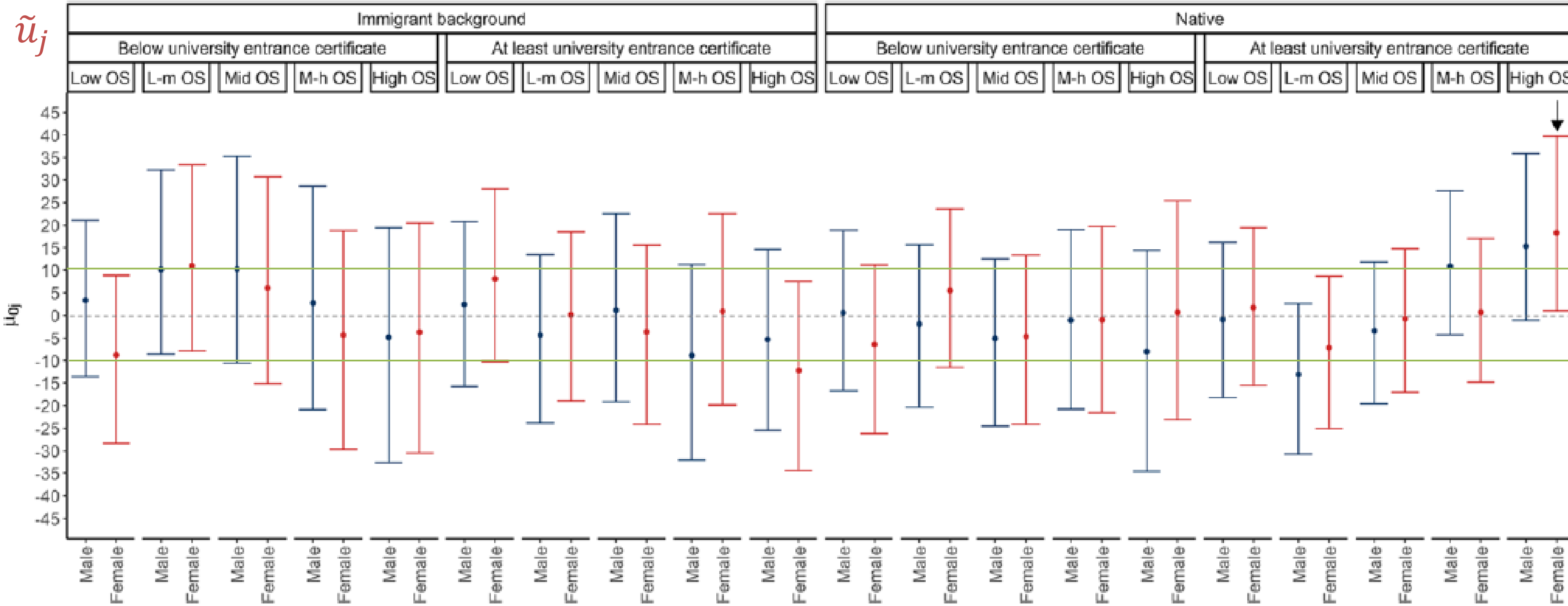
Table 3 Parameter estimates for the multilevel models of reading achievement in 15-year-old students

	Simple intersectional model estimates [95% CI]	Intersectional interaction model estimates [95% CI]
Fixed effects		
Intercept	478.26 [464.27, 492.26]	468.40 [452.81, 483.34]
Gender		
Female (reference)		–
Male		−25.62 [−36.46, −14.91]
Immigrant background		
Native (reference)		–
Immigrant background		−42.46 [−53.71, −30.90]
Highest parental education		
Below university entrance certificate (reference)		–
At least university entrance certificate		29.79 [18.23, 41.99]
Highest parental occupational status (HISEI)		
Low occupational status (reference)		–
Low to middle occupational status		10.06 [−5.85, 26.77]
Middle occupational status		27.63 [11.43, 44.58]
Middle to high occupational status		53.09 [35.19, 70.36]
High occupational status		62.59 [40.50, 81.23]
Measures of variance		
Between-stratum variance	1698.264	144.425
Within-stratum variance	9011.705	9021.200
VPC	15.86%	1.64%
PCV		91.15%

Note. 95% CI = 95% credible intervals; VPC = variance partition coefficient; PCV = proportional change in the between-strata variance

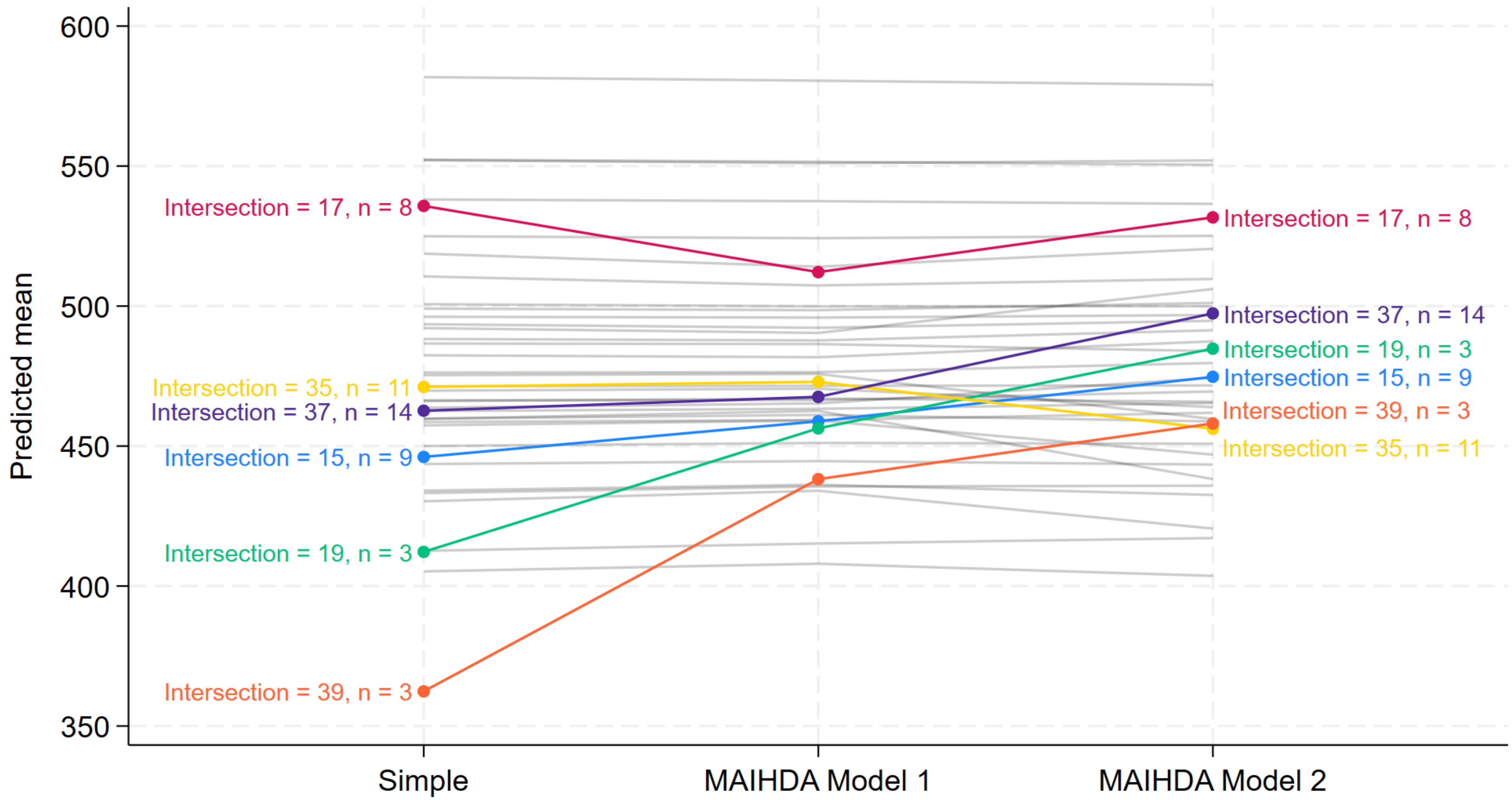
- **MAIHDA Model 2**
- Adds main effects of the four chars.
- $\widehat{\text{PCV}} = 0.91$
- 91% of intersectional inequalities are due to additive effects
- 9% to interactions (two-way and higher)
- Collectively, these interactions are statistically significant.

$\tilde{u}_j$



- Some of the intersection specific mean deviations from additivity are large:
  - Six are 10 points or higher (0.1 of a SD).
- However, **only one of these mean deviations is statistically significant**:
  - Native, female students with university educated, high occupation parents

But …. Simple, MAIHDA Model 1 and MAIHDA Model 2 means all differ!



- Differences especially large for small intersections.
- Why do we see differences, and which means are therefore best to report?

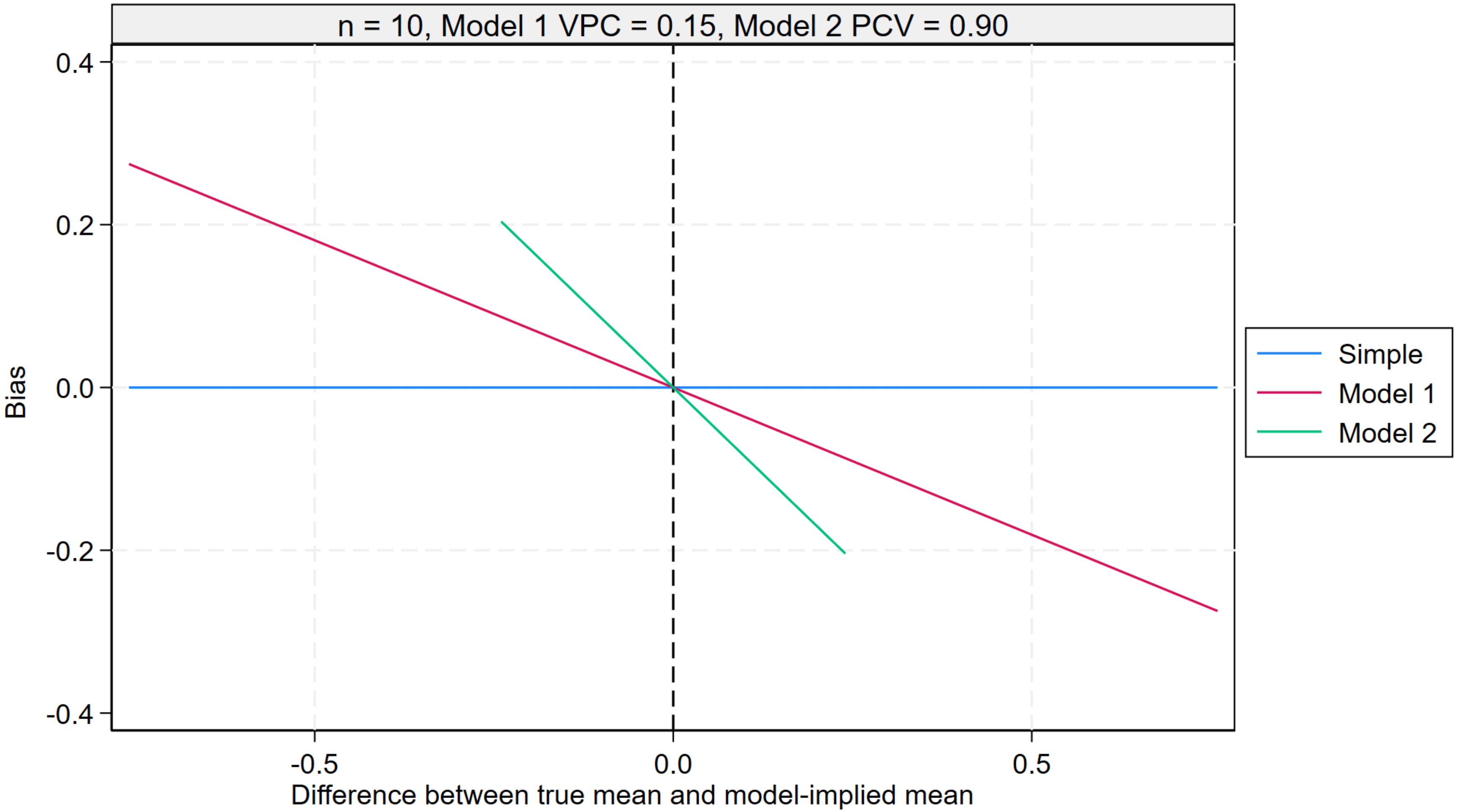
Analytic expressions for Bias, Variance, and MSE for a given intersection mean

- First, we derive analytic expressions for the …
  - **Bias of each mean**
  - **Variance of each mean**
  - **MSE of each mean (= Variance + bias^2)**

- Then, we plot these expressions against the deviation of the true means from the model-implied means (i.e., fixed part) assuming…
  - Model 1 VPC = 15%: *The importance of the inequalities*
  - Model 2 PCV = 90%: *The degree to which inequalities are additive*
  - $n_j = 10$: *Intersection size*

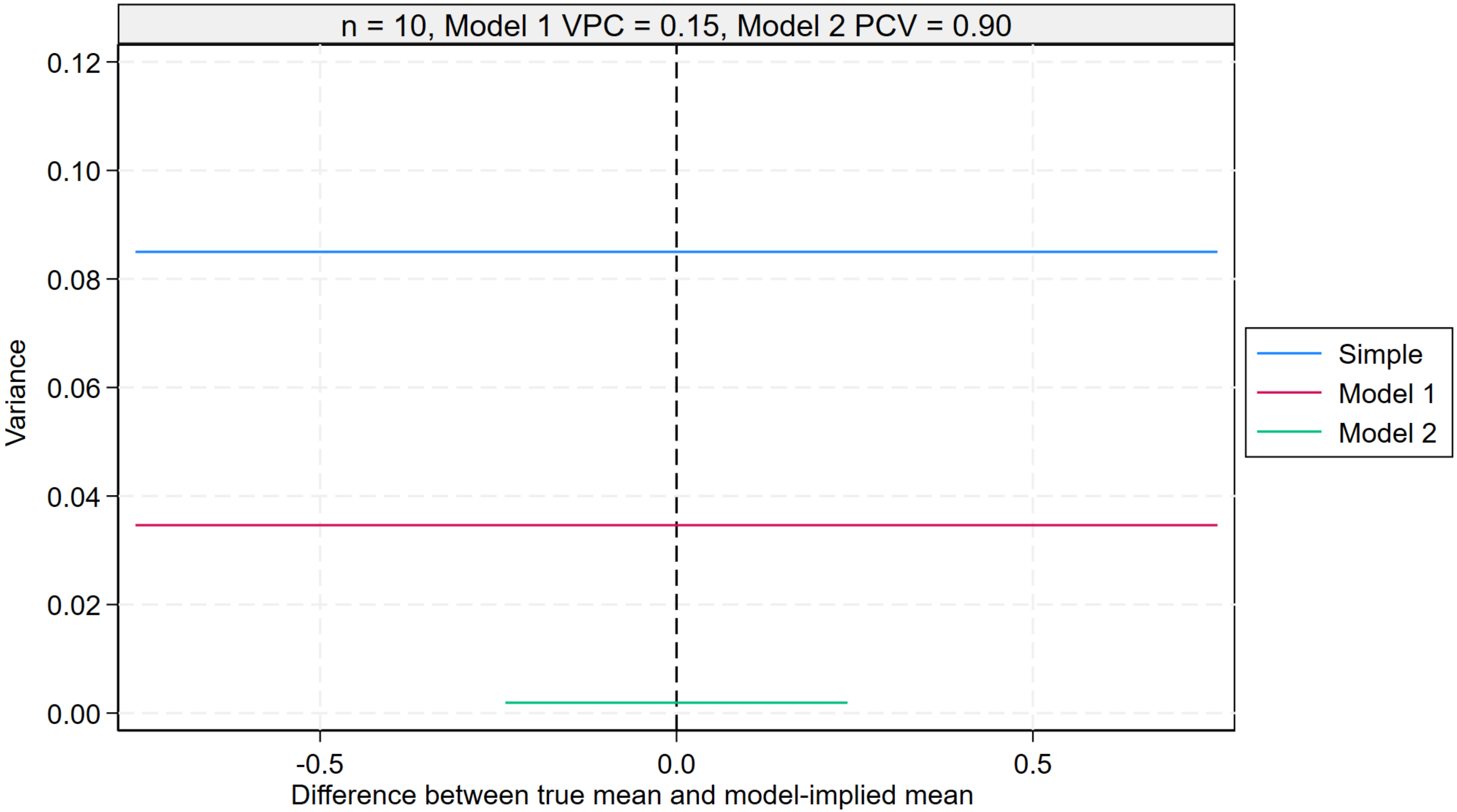
- To derive all expressions, we must declare a DGP:
  - We assume Model 2 is the true model

Bias for a specific intersection



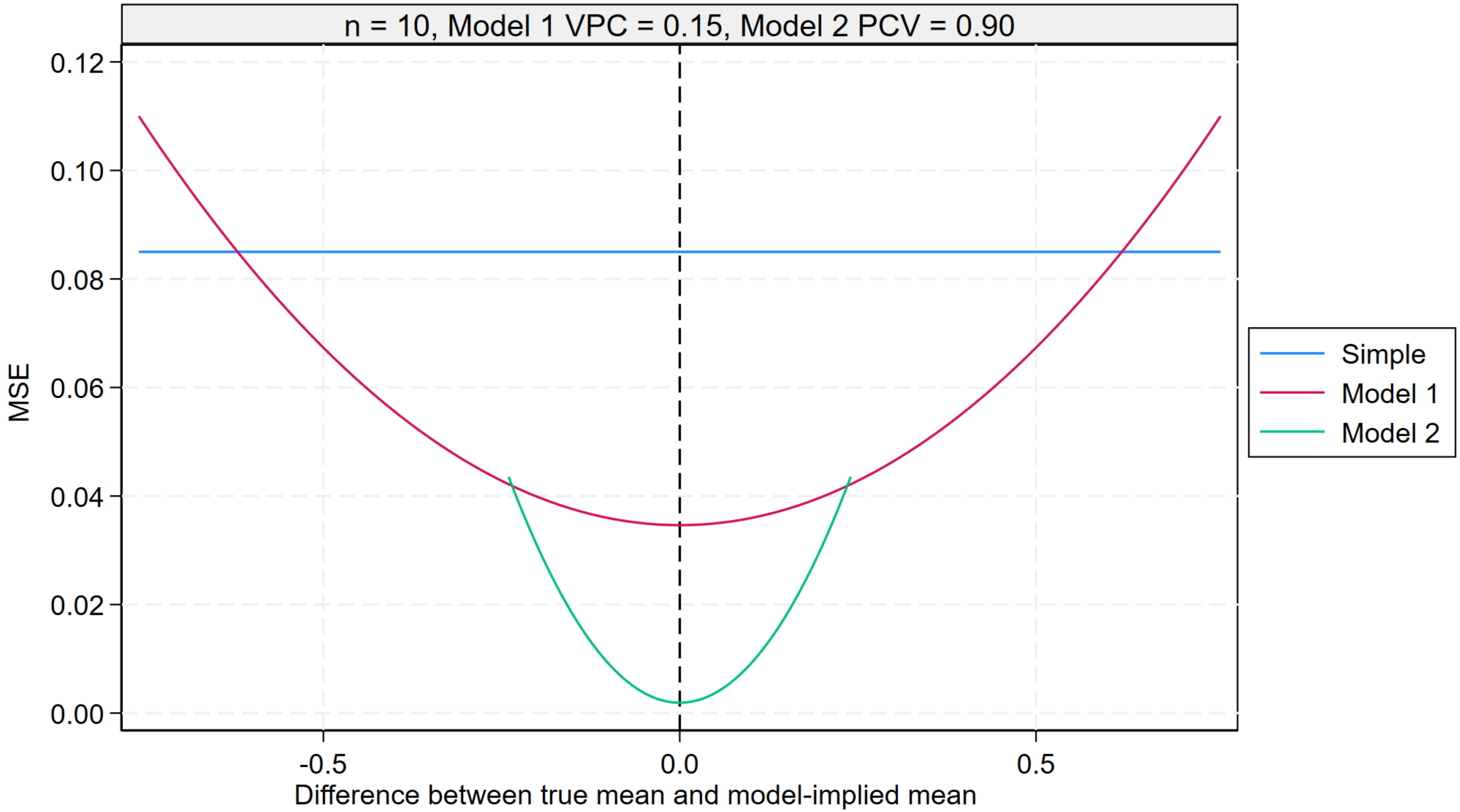
- S is unbiased. M1 biased towards grand mean of 0 as shrinkage is towards grand mean. M2 less biased as shrinkage is towards model-implied mean (additive).

Variance for a specific intersection



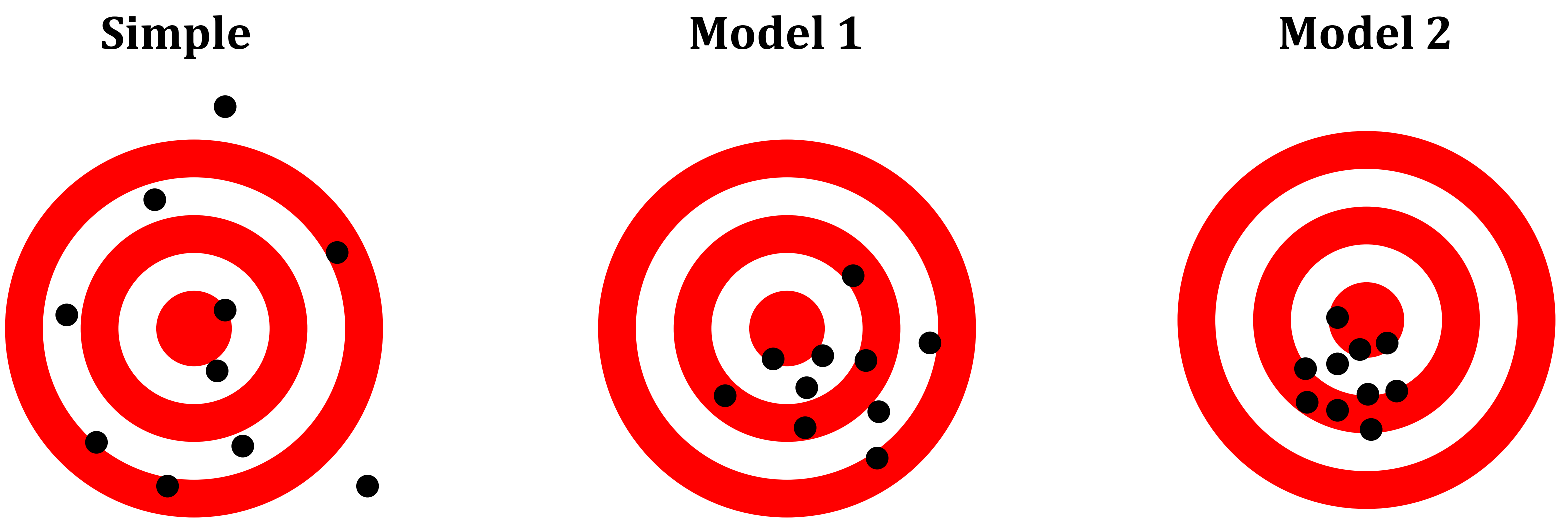
- S highest variance. M1 lower than S due to consistency in shrinkage across samples. M2 lower still due to greater shrinkage induced by high PCV.

MSE for a specific intersection



- S constant MSE as unbiased and var. constant. All other predictors quadratic. M2 lowest MSE.

Summary: The bias-variance trade-off



- Across repeated samples …
  - The simple mean will on average hit bull's eye (unbiased), but in any given sample it may miss the target entirely (large variance).
  - The Model 1 mean will on average miss the bull's eye (biased), but in any given sample it will not be so far off (smaller variance).
  - The Model 2 mean will on average get closer to the bull's eye (smaller bias) than the Model 1 mean and with more consistency (smallest variance).
- In any real analysis we only have one sample, so we go with Model 2.

SSM - Population Health 26 (2024) 101664

ELSEVIER

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

A tutorial for conducting intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA)

Clare R. Evans [a,*], George Leckie [b], S.V. Subramanian [c,d], Andrew Bell [e], Juan Merlo [f,g]

[a] *Department of Sociology, University of Oregon, Eugene, OR, USA*
[b] *Centre for Multilevel Modelling and School of Education, University of Bristol, UK*
[c] *Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA*
[d] *Harvard Center for Population and Development Studies, Cambridge, MA, USA*
[e] *Sheffield Methods Institute, University of Sheffield, Sheffield, UK*
[f] *Research Unit of Social Epidemiology, Faculty of Medicine, University of Lund, Sweden*
[g] *Center for Primary Health Care Research, Region Skåne, Malmö, Sweden*

A R T I C L E I N F O

Keywords:
MAIHDA
Intersectionality
Multilevel models
Health inequality
Linear regression
Logistic regression
Quantitative methods
Social determinants

A B S T R A C T

Intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (I-MAIHDA) is an innovative approach for investigating inequalities, including intersectional inequalities in health, disease, psychosocial, socioeconomic, and other outcomes. I-MAIHDA and related MAIHDA approaches have conceptual and methodological advantages over conventional single-level regression analysis. By enabling the study of inequalities produced by numerous interlocking systems of marginalization and oppression, and by addressing many of the limitations of studying interactions in conventional analyses, intersectional MAIHDA provides a valuable analytical tool in social epidemiology, health psychology, precision medicine and public health, environmental justice, and beyond. The approach allows for estimation of average differences between intersectional strata (stratum inequalities), in-depth exploration of interaction effects, as well as decomposition of the total individual variation (heterogeneity) in individual outcomes within and between strata.

Specific advice for conducting and interpreting MAIHDA models has been scattered across a burgeoning literature. We consolidate this knowledge into an accessible conceptual and applied tutorial for studying both continuous and binary individual outcomes. We emphasize I-MAIHDA in our illustration, however this tutorial is

The Statistical Advantages of MAIHDA for Estimating Intersectional Inequalities

George Leckie[1], Andy Bell[2], Juan Merlo[3], SV Subramanian[4], Clare Evans[5]

[1]Centre for Multilevel Modelling and School of Education, University of Bristol, UK.

[2]Sheffield Methods Institute, School of Education, University of Sheffield, Sheffield, UK

[3]Research Unit of Social Epidemiology, Faculty of Medicine, University of Lund, Sweden

[4]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[5]Department of Sociology, University of Oregon, Eugene, OR, USA

Contact information for the corresponding author

George Leckie

Centre for Multilevel Modelling and School of Education

University of Bristol

Two questions for the audience!

1. **Do you think MAIHDA is a valid application of multilevel modelling?**
   - From a theoretical perspective things look a little peculiar!
   - But, from a pragmatic perspective MAIHDA appears useful.

2. **What is the best way to demonstrate the greater predictive accuracy of multilevel model predicted intersection means over simple means?**
   - I have plotted analytic expressions.
   - But, I could have conducted a simulation study.

Thank you

End of talk