PREVALENCE AND STATISTICAL MODELING OF MULTIPLE MEMBERSHIP IN A STATEWIDE LONGITUDINAL DATA SYSTEM

Angela K. Henneberger

American Educational Research Association, Toronto

April 9, 2019

STRENGTHENING SOCIETY

UNIVERSITY of MARYLAND
SCHOOL OF SOCIAL WORK

CO-AUTHORS

- Yi Feng*
- Tessa Johnson*
- Yating Zheng*
- Bess A. Rose
- Laura M. Stapleton*
- Tracy Sweet*
- Michael E. Woolley
- *University of Maryland College Park

INTRODUCTION

- Statewide data systems are increasingly being used for research and policy evaluation (Figlio, 2017; Figlio et al., 2017).
- However, student mobility, where students attend multiple schools over time, presents a unique challenge to researchers.
  - 15-25% of students are mobile (U.S. GAO, 2002)
  - Mobility rates may be even higher for some student subgroups
- The traditional statistical approach used to model clustering is HLM (Raudenbush & Bryk, 2002).
  - However, data with mobile students violate the assumption that each student is nested within only one school.

STATISTICAL APPROACHES FOR HANDLING STUDENT MOBILITY

- HLM deleting mobile students
  - May reduce power
  - Limits external validity
- HLM assigning mobile students to their first (or last) school attended
  - May misattribute variance at student and school levels
  - Limits internal validity
- Multiple membership modeling (Beretvas, 2010)
  - Models the weighted effects of each school attended by the student
  - May lead to more accurate estimations of student and school variance components (Chung & Beretvas, 2012; Goldstein et al., 2007; Leckie, 2009).

MULTIPLE MEMBERSHIP MODELING

- Assumes school-level residuals are independent
  - However, we know that student movement among schools is not random
- Simulation studies are typically generated based on:
  - Low to moderate levels rates of student mobility (e.g., 10-20%)
  - Minimal to moderate ICC values (e.g., 5-25%)
  - Low to moderate cluster sample sizes (e.g., 50-100)
- Applied research on students and schools tends to have student mobility rates, ICC values, and cluster sample sizes that vary much more widely.

THE CURRENT STUDY

- The current study used statewide longitudinal data in Maryland to investigate:
  - The prevalence of multiple membership in statewide data
  - The prevalence of multiple membership for critical subgroups of students
  - The prevalence of multiple membership by school and district characteristics
  - Comparisons of the traditional HLM approach and multiple membership modeling

DATA AND COHORTS

- Data were from the Maryland Longitudinal Data System – Maryland's statewide repository for linked administrative educational and workforce data
- 9th grade students in 2009-2010
  - Enrolled in Maryland public school with grade span 9-12
  - Excluded exiters
  - (N = 61,364)

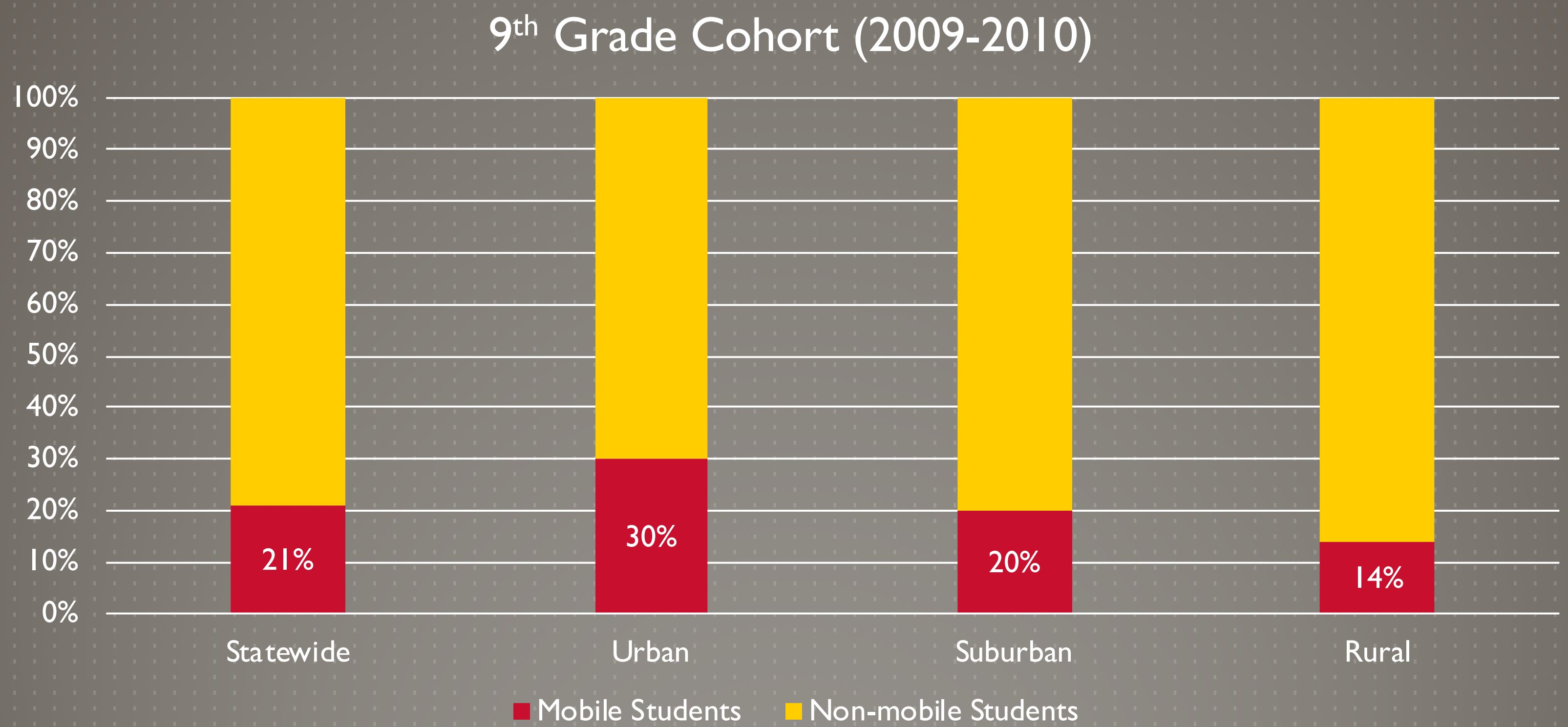
MEASURES

- Student mobility
  - Number of students who attended more than one school divided by total number of students in the cohort
  - *R* = 2-9 schools attended; 76% of mobile students attended 2 schools; 18% attended 3 schools; 5% attended 4 schools; 1% attended 5 or more schools
- Eligibility for free/reduced meals (FARMS; yes/no)
- Race/ethnicity (Hispanic; Black non-Hispanic; Other non-Hispanic; White)
- High school end of course assessments in Algebra and English (grand mean centered)

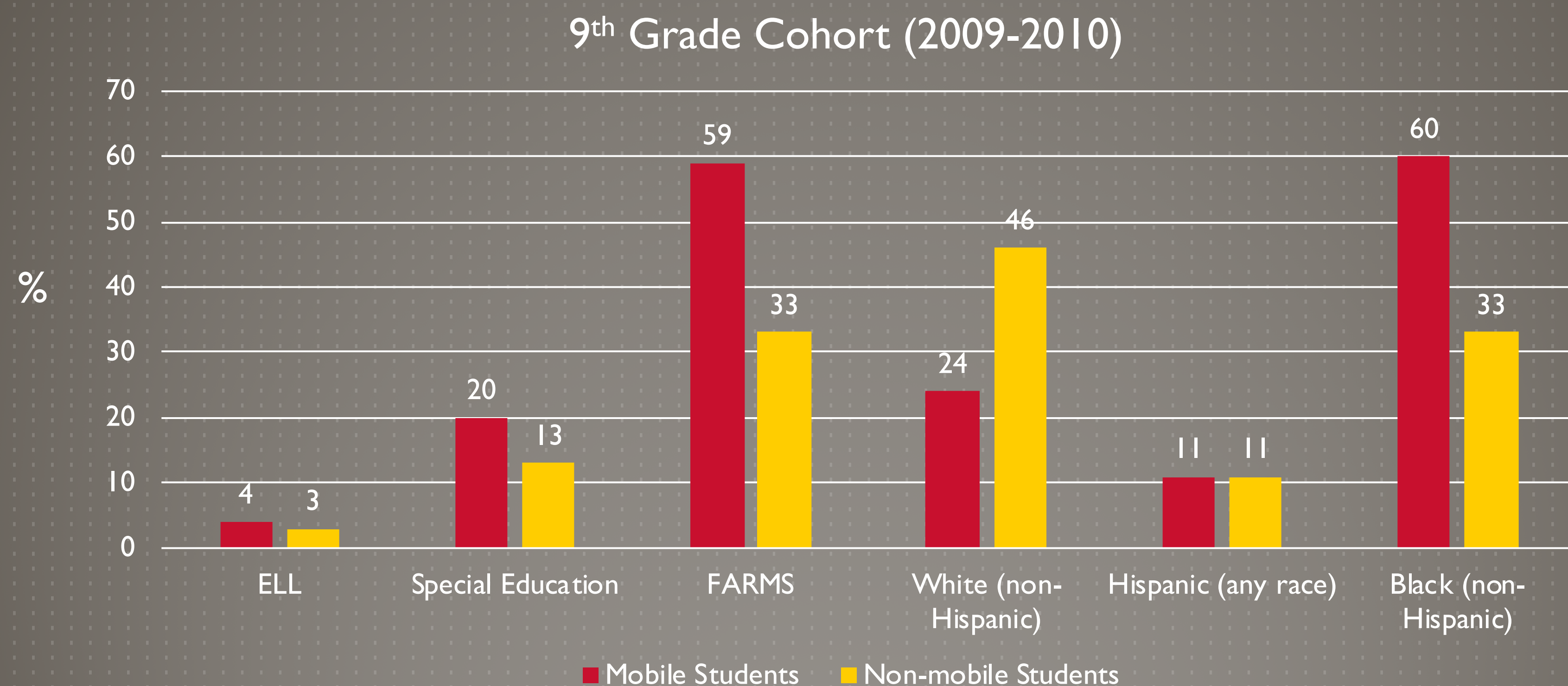
MEASURES (CONTINUED)

- College enrollment in the year following high school
  - Public and private enrollments
  - Maryland and out-of-state colleges
- Wages earned in the year following high school
  - For students who did not enroll in college
  - Log transformed

PREVALENCE OF MULTIPLE MEMBERSHIP



CHARACTERISTICS OF MOBILE STUDENTS



ANALYSES

Traditional HLM Approach

$Y_{ij} = \beta_{0j} + \beta_{1j}FARMS_{ij} + \beta_{2j}Hisp_{ij} + \beta_{3j}Black_{ij} + \beta_{4j}Other_{ij} + \beta_{5j}AlgHSA_{ij} + \beta_{6j}EngHSA_{ij} + e_{ij}$

$$\beta_{0j} = \gamma_{00} + \boxed{u_{0j}}$$

$$B_{1j} = \gamma_{10},\ B_{2j} = \gamma_{20},\ B_{3j} = \gamma_{30},\ B_{4j} = \gamma_{40},\ B_{5j} = \gamma_{50},\ B_{6j} = \gamma_{60}$$

Multiple Membership Approach

$Y_{i\{j\}} = \beta_{0\{j\}} + \beta_{1\{j\}}FARMS_{i\{j\}} + \beta_{2\{j\}}Hisp_{i\{j\}} + \beta_{3\{j\}}Black_{i\{j\}} + \beta_{4\{j\}}Other_{i\{j\}} + \beta_{5\{j\}}AlgHSA_{i\{j\}} + \beta_{6\{j\}}EngHSA_{i\{j\}} + e_{i\{j\}}$

$$\beta_{0\{j\}} = \gamma_{00} + \boxed{\sum_{h\in\{j\}} w_{ih}u_{0hj}}$$

$$B_{1\{j\}} = \gamma_{10},\ B_{2\{j\}} = \gamma_{20},\ B_{3\{j\}} = \gamma_{30},\ B_{4\{j\}} = \gamma_{40},\ B_{5\{j\}} = \gamma_{50},\ B_{6\{j\}} = \gamma_{60}$$

ICC CALCULATION

Intraclass Correlation Coefficient (ICC) Estimates for Multilevel Models Predicting College Enrollment and Wages

	College Enrollment			Wages		
	Model 1: HLM, Deletion approach	Model 2a: HLM, First-school approach	Model 2b: MM, MMREM approach	Model 1: HLM, Deletion approach	Model 2a: HLM, First-school approach	Model 2b: MM, MMREM approach
Statewide	0.00	0.00	0.87	0.00	0.00	0.05
Urban District	0.01	0.01	0.87	0.01	0.01	0.02
Suburban District	0.00	0.08	0.77	0.04	0.03	0.03
Rural District	0.18	0.32	0.80	0.04	0.03	0.08

Notes. ICC for college enrollment is calculated using the method proposed by Snijders & Bosker (1999) where the level 1 variance = Pj * (1 – Pj)

LOGISTIC MODEL PREDICTING COLLEGE ENROLLMENT

Fixed Effects and Variance Estimates for Logistic Multilevel Models Predicting College Enrollment (Statewide, 18.8% Mobility[e])

	Model 1: HLM, Deletion approach ($n_{students} = 49840$) ($N_{schools} = 221$)		**Model 2a: HLM, First-school approach** ($n_{students} = 61364$) ($N_{schools} = 273$)		**Model 2b: MM, MMREM approach** ($n_{students} = 61364$) ($N_{schools} = 285$)	
Fixed effects	γ_{0j}	$\exp(\gamma_{0j})$	γ_{0j}	$\exp(\gamma_{0j})$	γ_{0j}	$\exp(\gamma_{0j})$
Intercept	0.757 (0.016)	2.133	0.470 (0.016)	1.600	0.201 (0.051)	1.222
FARMS[a]	-0.603 (0.024)	0.547	-0.618 (0.021)	0.539	-0.513 (0.023)	0.599
Race/Eth: Hispanic vs. White NH[b]	0.238 (0.036)	1.269	0.238 (0.034)	1.268	0.087 (0.039)	1.091
Race/Eth: Black NH vs. White NH[b]	0.479 (0.026)	1.615	0.411 (0.024)	1.509	0.454 (0.031)	1.575
Race/Eth: Other NH vs. White NH[b]	0.796 (0.044)	2.216	0.810 (0.042)	2.248	0.666 (0.043)	1.946
HSA Algebra[c]	0.018 (0.001)	1.018	0.018 (0.001)	1.018	0.017 (0.001)	1.017
HSA English[c]	0.029 (0.001)	1.029	0.030 (0.001)	1.031	0.028 (0.001)	1.029
Random parameters						
Level-2 (Schools)						
Variance component, τ_{00}	0.000 (0.000)		0.000 (0.000)		0.670 (0.085)	
Level-1 (Students)						
Variance component, σ^2	---		---		---	
DIC	52090.73[d]		65703.49[d]		63695.22[d]	

Notes. Standard errors are in parentheses. Analysis conducted for the cohort of students who were in 9th grade in 2009-2010.
[a] A binary variable indicating eligibility to receive free/reduced price meals in school year 2009-2010 (0=Not eligible, 1=Eligible).
[b] A set of dummy variables indicating race/ethnicity: Hispanic, Black non-Hispanic, Other non-Hispanic (with White non-Hispanic as the reference group).
[c] HSA = High School Assessment. HSA assessment scores are all grand-mean centered.
[d] DIC is only comparable for Models 2a & 2b. The DIC for Model 1 cannot be compared due to differing student sample sizes.
[e] Proportion of students attending more than one school.

MODEL PREDICTING WAGES (LOG TRANSFORMED)

Fixed Effects and Variance Estimates for Multilevel Models Predicting Log-transformed Wages (Statewide, 23.1% Mobility[e]*)*

	Model 1: HLM, Deletion approach ($n_{students}$ = 7071) ($N_{schools}$ = 207)		Model 2a: HLM, First-school approach ($n_{students}$ = 9273) ($N_{schools}$ = 253)		Model 2b: MM, MMREM approach ($n_{students}$ = 9273) ($N_{schools}$ = 264)	
Fixed effects	γ_{0j}	$\exp(\gamma_{0j})$	γ_{0j}	$\exp(\gamma_{0j})$	γ_{0j}	$\exp(\gamma_{0j})$
Intercept	8.624 (0.023)	5563.596	8.611 (0.021)	5491.74	8.539 (0.028)	5110.232
FARMS[a]	-0.056 (0.033)	0.946	-0.061 (0.028)	0.941	-0.069 (0.029)	0.933
Race/Eth: Hispanic vs. White NH[b]	0.030 (0.056)	1.030	0.080 (0.051)	1.083	0.172 (0.056)	1.188
Race/Eth: Black NH vs. White NH[b]	-0.383 (0.036)	0.682	-0.380 (0.031)	0.684	-0.292 (0.037)	0.747
Race/Eth: Other NH vs. White NH[b]	-0.198 (0.072)	0.820	-0.181 (0.064)	0.835	-0.115 (0.065)	0.891
HSA Algebra[c]	0.003 (0.001)	1.003	0.003 (0.001)	1.003	0.003 (0.001)	1.003
HSA English[c]	-0.007 (0.001)	0.993	-0.006 (0.001)	0.994	-0.006 (0.001)	0.994
Random parameters						
Level-2 (Schools)						
Variance component, τ_{00}	0.001 (0.001)		0.000 (0.000)		0.049 (0.009)	
Level-1 (Students)						
Variance component, σ^2	1.553 (0.026)		1.562 (0.023)		1.524 (0.022)	
DIC	23192.58[d]		30461.24[d]		30338.09[d]	

Notes. Standard errors are in parentheses. HSA = High School Assessment. Analysis conducted for students who were in 9th grade in 2009-2010, but did not enroll in postsecondary institutions in 2013-2014.

[a] A binary variable indicating eligibility to receive free/reduced price meals in school year 2009-2010 (0=Not eligible, 1=Eligible).

[b] A set of dummy variables indicating race/ethnicity: Hispanic, Black non-Hispanic, Other non-Hispanic (with White non-Hispanic as the reference group).

[c] HSA = High School Assessment. HSA assessment scores are all grand-mean centered.

[d] DIC is only comparable for Models 2a & 2b. The DIC for Model 1 cannot be compared due to differing student sample sizes.

[e] Proportion of students attending more than one school.

SUMMARY OF RESULTS

- The prevalence of multiple membership was high.
- Minority students and students living in poverty were disproportionately likely to be mobile.
- Highest mobility rate was in the urban school district.
- Large variations in the estimated ICC, intercept, fixed effects, and variance estimates.
- Largest variation in the urban and rural school districts, where mobility rates were high.

DISCUSSION

- Choice of modeling approach may define the plausibility of finding school effects
- HLM deletion approach may reduce statistical power and limit generalizability
  - Disproportionately delete minority students and students living in poverty
  - May lead to underestimation of relations with outcomes
- HLM first-school approach may misattribute school variance to the student level
  - May lead to overestimation of relation between student characteristics and outcomes, especially when student characteristic is highly correlated with school membership

DISCUSSION

- MMREM may more accurately attribute student and school level variance when compared to the other approaches
  - Must consider data available (e.g., districtwide data; statewide data; national data)
  - Introduced more clusters with only a few students nested within each cluster
  - MMREM is a critical tool for applied researchers to know about at the start of the study
- Inform future simulation studies examining MMREM
  - High mobility rates
  - High ICC
  - Low cluster sample size

ACKNOWLEDGEMENT

This manuscript was developed under a grant from the Department of Education. However, these contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government. The authors are grateful for the data, technical, and research support provided by the Maryland Longitudinal Data System (MLDS) Center and its agency partners. The views and opinions expressed are those of the authors and do not necessarily represent the views of the MLDS Center or its agency partners.

CONTACT

Dr. Angela Henneberger

University of Maryland School of Social Work

ahenneberger@ssw.umaryland.edu