MLDS Center

Maryland Longitudinal Data System

Better Data • Informed Choices • Improved Results

Multiple Membership Modeling Versus Traditional Multilevel Modeling for Handling Student Mobility in Maryland

Angela K. Henneberger
Bess Rose
MLDS Center & UMB SSW

MLDS Center Research Series
October 3, 2019

MLDS Center
Maryland Longitudinal Data System

Acknowledgements

Co-Authors: Yi Feng*; Tessa Johnson*; Yating Zheng*; Laura M. Stapleton*; Tracy Sweet*; Michael E. Woolley

*University of Maryland College Park

The contents of this presentation were 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.

MLDS Center
Maryland Longitudinal Data System

Overview

- Introduction and background
- Methods
- Results
- Discussion
- Questions

MLDS Center
Maryland Longitudinal Data System

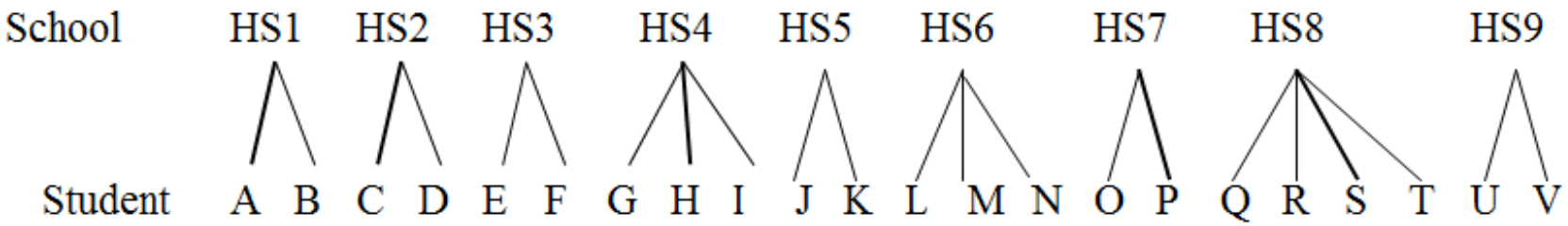
Introduction and Background

Introduction

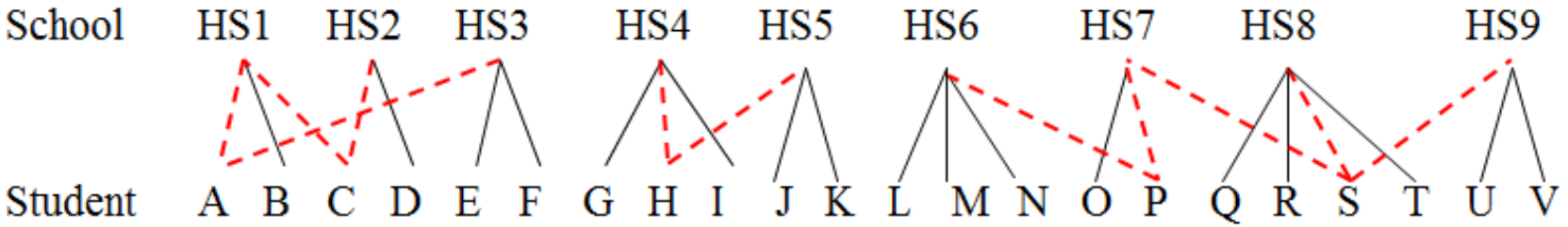
- Researchers using MLDS data are interested in student-level and/or school-level effects
  - Student level: e.g. participation in CTE
  - School level: e.g. concentration of poverty
- Education data are inherently clustered (students are nested within classrooms, which are nested within schools, which are nested within districts)
- Analyzing one level without the other will produce misleading results
- Hierarchical Linear Modeling (HLM; Raudenbush & Bryk, 2002) is the traditional statistical approach for correctly adjusting for clustering

Pure Hierarchical Design

HLM is appropriate when each student is nested within *only one* school (HS=high school):



… but real-world data aren't purely hierarchical!



MLDS Center
Maryland Longitudinal Data System

The Statistical Problem: Student Mobility

- Student mobility - when students change schools either within the academic year or between academic years (Rumberger, 2002)
- Mobility types: out of school, out of district, out of state/out of public schools
- Mobility rates are high in the United States (U.S. Government Accounting Office (1994):
  - 15% of suburban 1st-3rd grade students are mobile
  - 25% of urban 1st-3rd grade students are mobile
  - Rates are even higher for some student subgroups

MLDS Center
Maryland Longitudinal Data System

The Statistical Problem: Student Mobility (cont'd)

- Longitudinal data present an additional problem because students are more likely to attend multiple schools with each additional year of data
- Remember, HLM assumes that each student is nested within only one school
- Assumption is violated when students are mobile

Solutions?

- Common but problematic solutions:
  - Deleting mobile students
    - Reduces statistical power
    - Limits generalizability (external validity)
  - Assigning mobile students to their first school attended
    - Leads to misattributing student and school level variance
    - Limits internal validity
- Better solution: multiple membership multilevel modeling

(Beretvas, 2010; Chung, 2009; Chung & Beretvas, 2012; Goldstein, Burgess & McConnell, 2007)

Multiple Membership Modeling

Traditional HLM:

Student level: $Y_{ij} = \beta_{0j} + e_{ij}$

School level: $\beta_{0j} = \gamma_{00} + u_{0j}$

Multiple Membership Model: Model the weighted effects of each school attended by each student

Student level: $Y_{i\{j\}} = \beta_{0\{j\}} + e_{i\{j\}}$

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

The Current Study

(1) What is the prevalence of mobility for students in Maryland?
   (a) How common are the different types of mobility (out of schools, out of districts, and out of Maryland public schools)?
   (b) How much mobility occurs after 1, 2, 3 years?
(2) What is the prevalence of mobility for specific subgroups of students and types of schools?
(3) How do results differ for two statistical approaches for handling student mobility (traditional multilevel modeling vs. multiple membership modeling)?

Methods

Methods: Overall Prevalence

1. What is the prevalence of mobility for students in Maryland?

→ For representative cohorts (6th grade, 9th grade), compare original school to school 1, 2, 3 years later

→ Produce descriptive statistics (count and percentage) for each cohort by mobility type and year

MLDS Center
Maryland Longitudinal Data System

Overall Prevalence Cohorts

- Middle School (MS) Cohort: 6th grade students (2014-2015; N = 60,062)
- High School (HS) Cohort: 9th grade students (2013-2014; N = 71,555)

Methods: Prevalence by Groups

2. What is the prevalence of mobility for specific subgroups of students and types of schools?

→ Produce descriptive statistics (counts and percentages) for each subgroup and type

Methods: Prevalence by Group Characteristics

- Student-Level
  - MS & HS: English Learner (EL)
  - MS & HS: Special Education
- School-Level
  - School poverty: % of students eligible for free/reduced price meals (FARMS; low/medium/high)
- Measured in the year of initial enrollment or the year prior

Methods: Model Comparisons

3. How do results differ when using traditional multilevel modeling versus multiple membership modeling with mobile students?

→ Traditional HLM model - delete mobile students

→ Traditional HLM model - assign mobile students to their first school

→ Multiple membership (MM) model

→ Compare the results

Methods: Model Comparisons Cohort

- 9th grade students in 2009-2010
  - Enrolled in Maryland public school with grade span 9-12
  - Excluded exiters
  - (N = 61,364)

MLDS Center
Maryland Longitudinal Data System

Methods: Model Comparisons Covariates

- 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)

Methods: Model Comparisons Outcomes

- 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

Methods: Model Comparisons Analyses

Traditional HLM Approach

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

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

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

Multiple Membership Approach

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

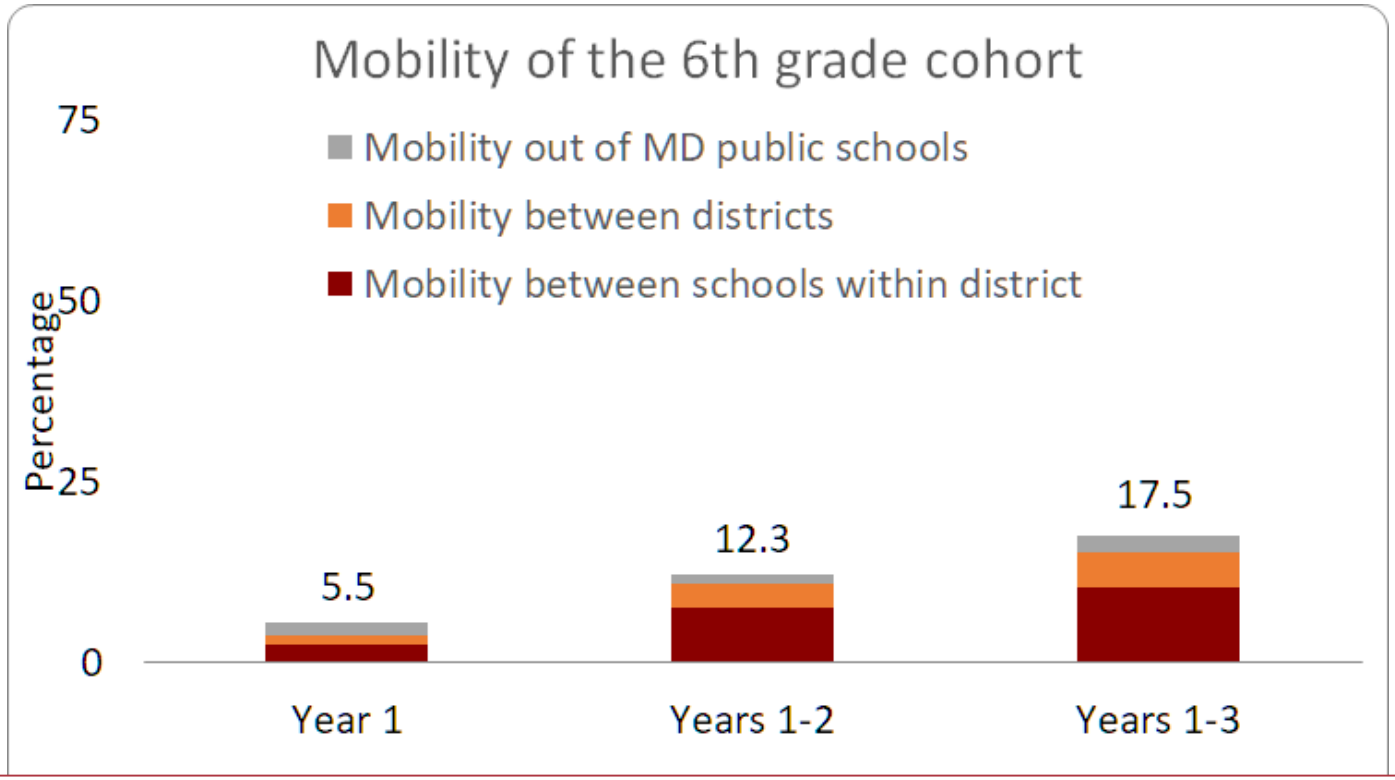
$$\beta_{0\{j\}} = \gamma_{00} + \sum_{h \in \{j\}} w_{ih} u_{0hj}$$

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

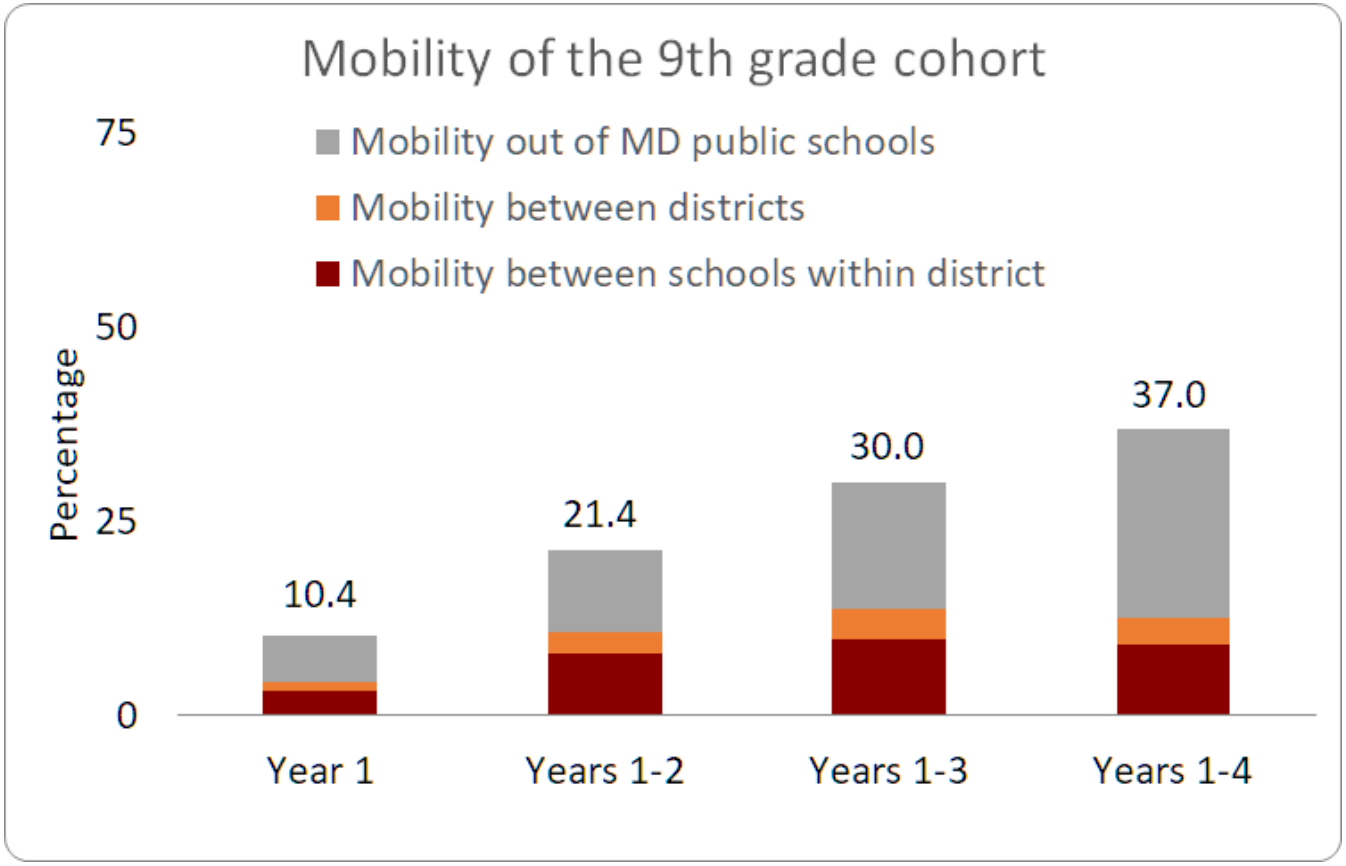
Results

Results: Overall Prevalence

1. What is the prevalence of mobility for students in Maryland?
    (a) How common are the different types of mobility (out of schools, out of districts, and out of Maryland public schools)?
    (b) How much mobility occurs after 1, 2, 3 years?



- Within the 1st year, 5.5 percent of the 6th grade cohort moved out of the school where they had started 6th grade.
- By the end of the 2nd year, 12.3 percent had moved.
- By the end of 3 years, 17.5 percent had moved, most staying within the same district (10.4%). 4.9% had moved between districts. 2.3% had moved out of MD public schools.



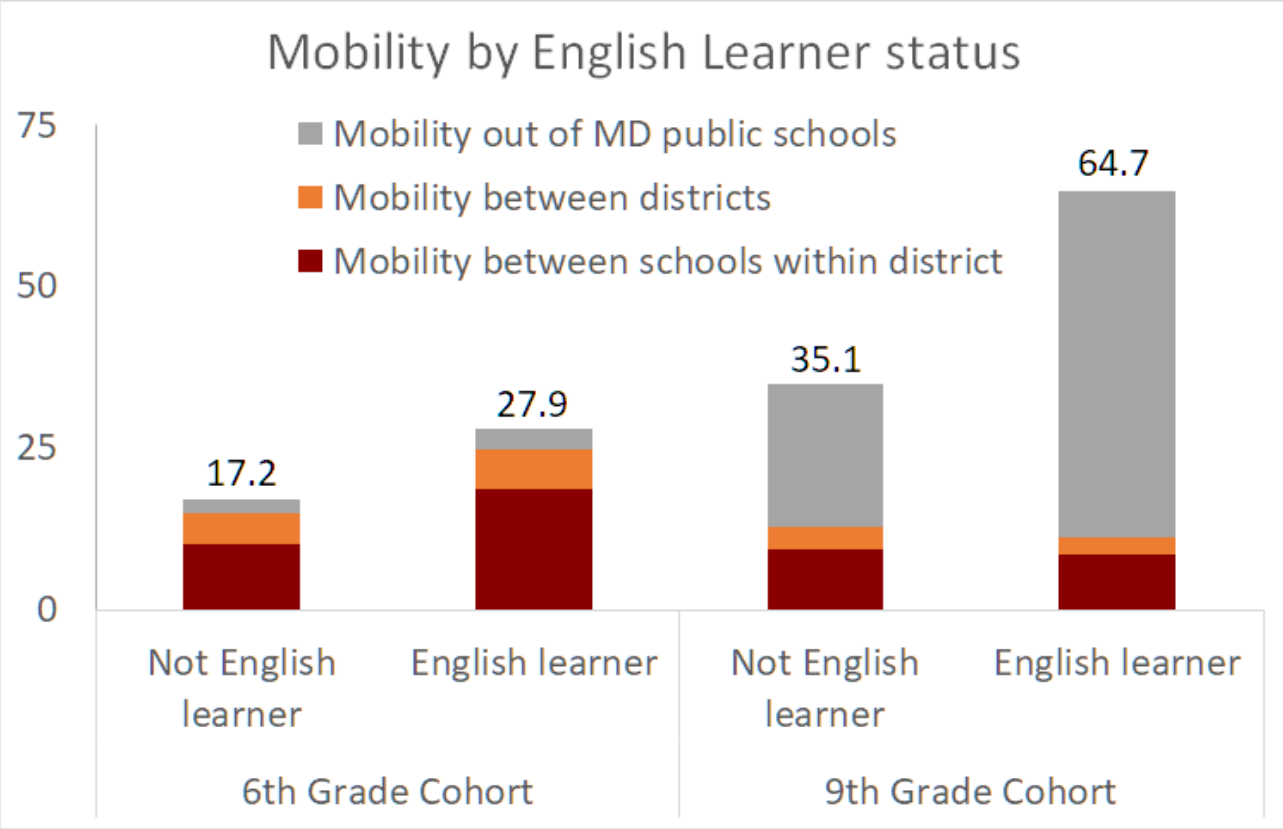
- By the end of 4 years, 37% of the 9th grade cohort experienced mobility out of the school where they started 9th grade.
- Most of this mobility was out of MD public schools altogether.

Summary: Overall Prevalence

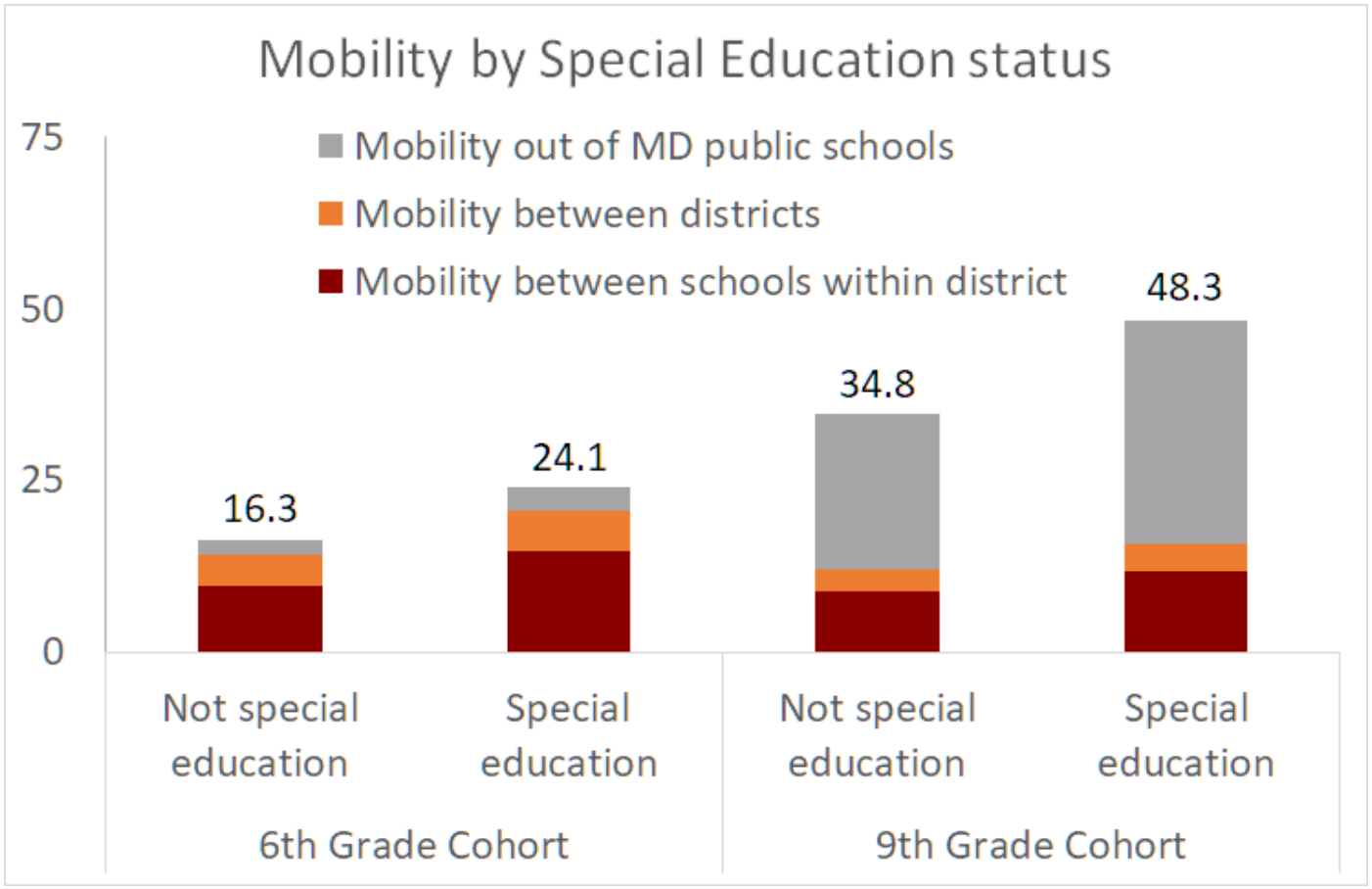
- Mobility accumulated each year
- Higher mobility among high schoolers
- Accounting for school clustering would be problematic for a strictly hierarchical model
- Particularly for the high school cohort, there are a sizeable number of transfers out of MD public schools - a problem for our longitudinal, cross-sector studies
  - We won't know whether they graduate from HS
  - We won't know whether they enroll in college out of state (National Student Clearinghouse data are only obtained for MD HS graduates)

Results: Prevalence by Groups

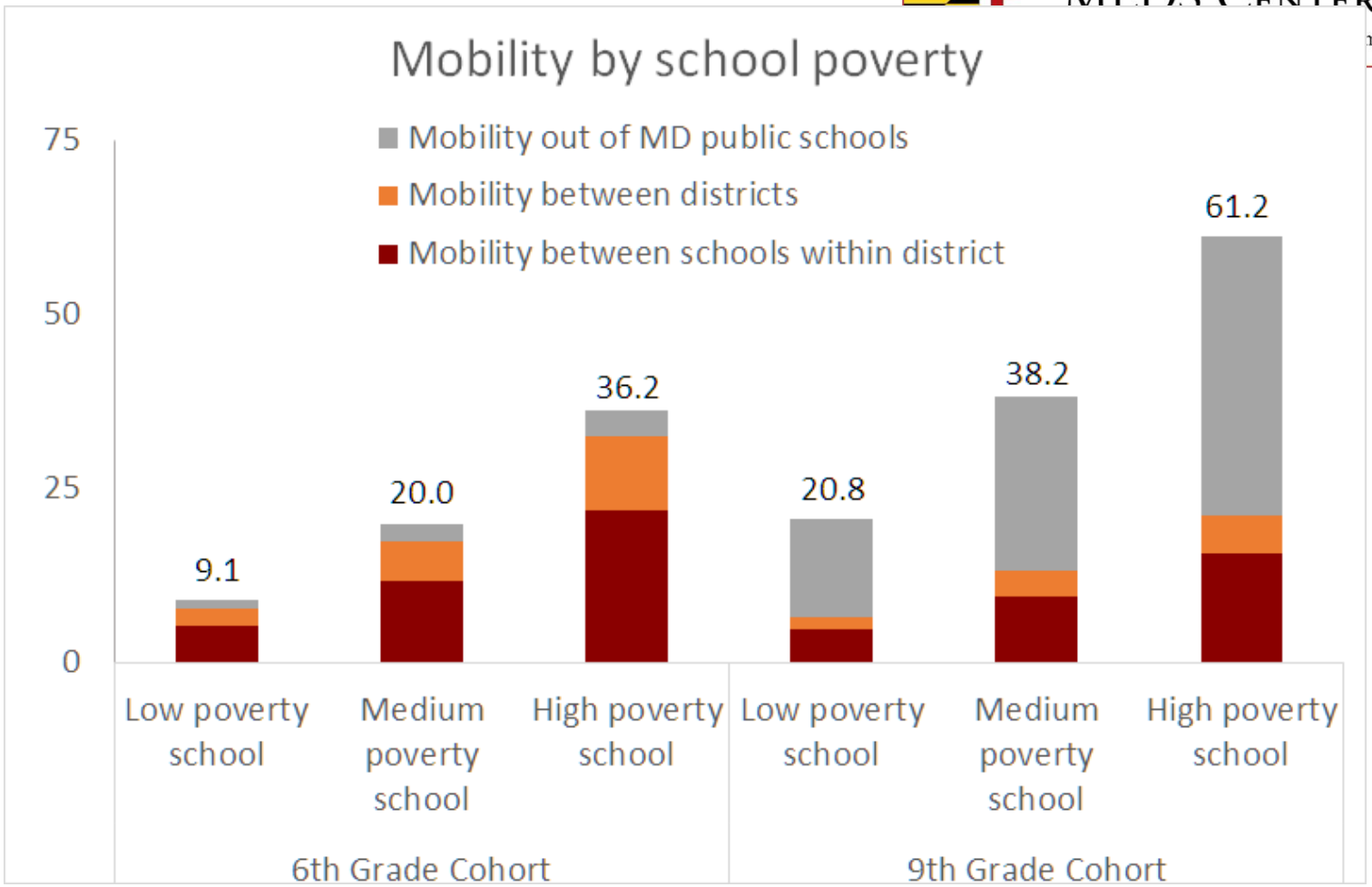
2. What is the prevalence of mobility for specific subgroups of students and types of schools?



- Mobility rates were higher for English learners than for non-ELs.
- For the 6th grade cohort, the types of mobility were comparable. For the 9th grade cohort, ELs were more likely to transfer out of MD public schools altogether.



- Mobility rates were higher for special education students.
- For both cohorts, the types of mobility were comparable between students receiving and not receiving special education services.



- Mobility rates increased as school poverty increased.
- For both cohorts, the types of mobility were comparable across school poverty levels. Most mobility among high schoolers is out of MD public schools - in both low poverty and high poverty schools.

Summary: Prevalence by Groups

- Higher mobility for ELs and special education students
  - Incorrectly accounting for clustering could be particularly problematic for inferences about these students
- Higher mobility in higher poverty schools
  - Incorrectly accounting for clustering could be particularly problematic for inferences about these schools
- Transfer out of MD public schools is much higher for ELs than for non-ELs in high school
  - This “differential attrition” is particularly problematic for valid inferences

Results: Model Comparisons

3. How do results differ when using traditional multilevel modeling versus multiple membership modeling with mobile students?

→ Traditional HLM model - delete mobile students

→ Traditional HLM model - assign mobile students to their first school

→ Multiple membership (MM) model

→ Compare the results

Multilevel model results: Log wages	**Model 1: HLM (Delete)**	**Model 2: HLM (Use first school)**	**Model 3: Multiple membership**
Student N=9,273 School N=264	Student N=7,071 School N=207	Student N=9,273 School N=253	Student N=9,273 School N=264
Intercept	8.624 (0.023)	8.611 (0.021)	8.539 (0.028)
FARMS	-0.056 (0.033)	-0.061 (0.028)	-0.069 (0.029)
Hispanic	0.030 (0.056)	0.080 (0.051)	0.172 (0.056)
Black	-0.383 (0.036)	-0.380 (0.031)	-0.292 (0.037)
Other race/ethnicity	-0.198 (0.072)	-0.181 (0.064)	-0.115 (0.065)
HSA Algebra	0.003 (0.001)	0.003 (0.001)	0.003 (0.001)
HSA English	-0.007 (0.001)	-0.006 (0.001)	-0.006 (0.001)
Level 2 (schools) variance	0.001 (0.001)	0.000 (0.000)	0.049 (0.009)
Level 1 (students) variance	1.553 (0.026)	1.562 (0.023)	1.524 (0.022)
DIC†	23192.58†	30461.24†	30338.09†

† DIC is only comparable for Models 2 & 3. The DIC for Model 1 cannot be compared due to differing student sample sizes.

Multilevel model results: Likelihood of college enrollment	**Model 1: HLM (Delete)**	**Model 2: HLM (Use first school)**	**Model 3: Multiple membership**
Student N=61,364 School N=285	Student N=49,840 School N=221	Student N=61,364 School N=273	Student N=61,364 School N=285
Intercept	0.757 (0.016)	0.470 (0.016)	0.201 (0.051)
FARMS	-0.603 (0.024)	-0.618 (0.021)	-0.513 (0.023)
Hispanic	0.238 (0.036)	0.238 (0.034)	0.087 (0.039)
Black	0.479 (0.026)	0.411 (0.024)	0.454 (0.031)
Other race/ethnicity	0.796 (0.044)	0.810 (0.042)	0.666 (0.043)
HSA Algebra	0.018 (0.001)	0.018 (0.001)	0.017 (0.001)
HSA English	0.029 (0.001)	0.030 (0.001)	0.028 (0.001)
Level 2 (schools) variance	0.000 (0.000)	0.000 (0.000)	0.670 (0.085)
Level 1 (students) variance	--	--	--
DIC	52090.73†	65703.49†	63695.22†

† DIC is only comparable for Models 2 & 3. The DIC for Model 1 cannot be compared due to differing student sample sizes.

Summary: Model Comparisons

- Deleting mobile students results in losses of students and schools
  - 2,202 students (24%) and 57 schools (22%) lost for wage analysis
  - 11,524 students (19%) and 64 schools (22%) lost for college enrollment analysis
- Assigning mobile students to their first school results in losses of schools
  - 11 schools lost for wage analysis (4%)
  - 12 schools lost for college enrollment analysis (4%)

MLDS Center
Maryland Longitudinal Data System

Summary: Model Comparisons (cont'd)

- Estimates of student-level effects (coefficients) and their statistical significance (standard errors) vary considerably across models
- Proportion of variance attributable to differences between schools is underestimated by traditional purely hierarchical models
- Model fit statistics indicate multiple membership models are better than the first-school models

Discussion

Discussion

- The loss of students and schools when ignoring student mobility results in threats to external validity
  - Deleting mobile students results in disproportionate losses of some types of students (EL, minority, FARMS)
- 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, cont'd

- Multiple membership models 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
- Multiple membership modeling is a critical tool for applied researchers to know about at the start of the study

Limitations

- Limited understanding of students who leave the Maryland public school system
- Workforce wages are limited to individuals employed at employers subject to Maryland Unemployment Insurance
- Limited school-level variance in our currently examined outcomes - *future research*
- No inclusion of classroom-level variance - *future research*
- The supports and barriers to use of multiple membership modeling are unknown

Future Research

- How does choice of modeling approach affect estimation of school-level covariates?
  - Currently running models estimating effects for outcomes with larger school-level variance (e.g., SAT scores; PSAT scores; HSA scores)
- To what extent does having statewide (population) data alleviate the negative effects of not accounting for mobility in modeling approach?
- To what extent does mobility occur at the classroom and teacher levels?
  - Clustering at the classroom/teacher level
  - Modeling effects for teachers

MLDS Center
Maryland Longitudinal Data System

Questions and Contact

Dr. Angela Henneberger
MLDS Center Director of Research
angela.henneberger@maryland.gov

Dr. Bess A. Rose
MLDS Center Statistician
bess.rose@maryland.gov