Peer Effects and Differential Attrition

Evidence from Tennessee's Project STAR

Sanjay Satish

Duke University

November 29, 2021

Overview

1. Motivation & Thesis Overview
2. Identifying Attrition
3. Quantifying the Effect of Ability on Attrition
4. Identification of Differential Peer Effects From Attrition Across Class Size
5. Identification of Peer Effects From Attrition to Private Schools

Motivation

- Student-to-student peer effects do in fact exist and such effects are potentially influenced by peer quality
- Private schools (charters as well, although less important for my thesis) have entrance exam requirements that result in **cream-skimming**
- In schools, there perhaps exists evidence of **differential attrition** based on peer ability and subsequent **negative peer effects** from such attrition

Project STAR

- Class-size randomizatino expiriment across grades K-3 in TN between 1985-1989
- 11,600 students across 79 schools voluntairly participated in the study
- Students and teachers ranomized into one of 3 class types at beginning of 1985 school year
  1. Small Classes (13-17 students)
  2. Regular-sized classes (22-25 students)
  3. Regular-sized classes with teacher's aide (22-25 students + full-time teacher's aide)

Project STAR

- After being assigned to a class type, students kept in that class for duration of expiriment
- Each participating school required to have one of three class types and randomization occourred within each school
- Key limitation is that at the beginning of 1st grade, students in regular and regular-aide classes were re-randomized across class types
- $\approx$ 10% of students moved across small and regular-sized classes between grades

Summary Stats - Grade 1

	Small Class	Regular Class	Regular Class with Aide
Experimental Characteristics			
Number of Years in STAR	3.2 ± 1.0	2.9 ± 1.1	3.0 ± 1.0
Number of Years in Small Classes	3.0 ± 1.0	0.2 ± 0.5	0.1 ± 0.3
School Urbanicity			
Inner City	19.8%	22.1%	18.5%
Suburban	23.5%	24.9%	21.2%
Rural	47.3%	45.0%	50.1%
Urban	9.4%	8.0%	10.2%
Teacher Characteristics			
Female	97.5%	100.0%	100.0%
White	81.2%	83.8%	81.6%
Years of Experience	12.2 ± 8.7	10.3 ± 8.7	12.7 ± 9.2
Class Size	15.7 ± 1.6	22.7 ± 2.3	23.4 ± 2.4
Receives Free Lunch	47.8%	51.9%	50.3%
Special Education	0.6%	1.5%	1.5%
Math SAT Scaled Score	538.7 ± 44.1	525.3 ± 41.7	529.6 ± 42.9
Reading SAT Scaled Score	530.0 ± 56.6	513.5 ± 53.5	521.3 ± 54.7
Listening SAT Scaled Score	572.7 ± 34.5	563.8 ± 32.4	567.2 ± 33.9
Word Study Skills SAT Scaled Score	523.0 ± 52.6	506.2 ± 54.0	513.7 ± 51.8

Note: some variables expressed as Mean ± Standard Deviation

Research Question(s)

Essential Question(s):

What was the nature of attrition in Project STAR? Secondly, if student ability was a significant factor in attrition, what are the peer effects from such attrition?

Classifying Attrition

- There are many new papers on the long-run impacts of Project STAR
- As part of such papers, data has been collected on academic achievment of Project STAR students in grades 4 through 8
- Utilizing this data, it is possible to identify students who left the public school system during the expiriment through matching

Reproducibility

- Rohlfs and Zilora (2014) identify 5 types of attrition within Project STAR
  1. Students who were in a Project STAR school at time $t-1$ but who's school left the expiriment at time t
  2. Students who changed class type within a school
  3. Students who left to another Project STAR school
  4. Students who left to another public school
  5. Students who left the public school system
- I aim to reproduce Rohlfs and Zilora's results, specifically strengthening identification by utilizing testing records for both the 4th and 5th grade

Attrition Chart

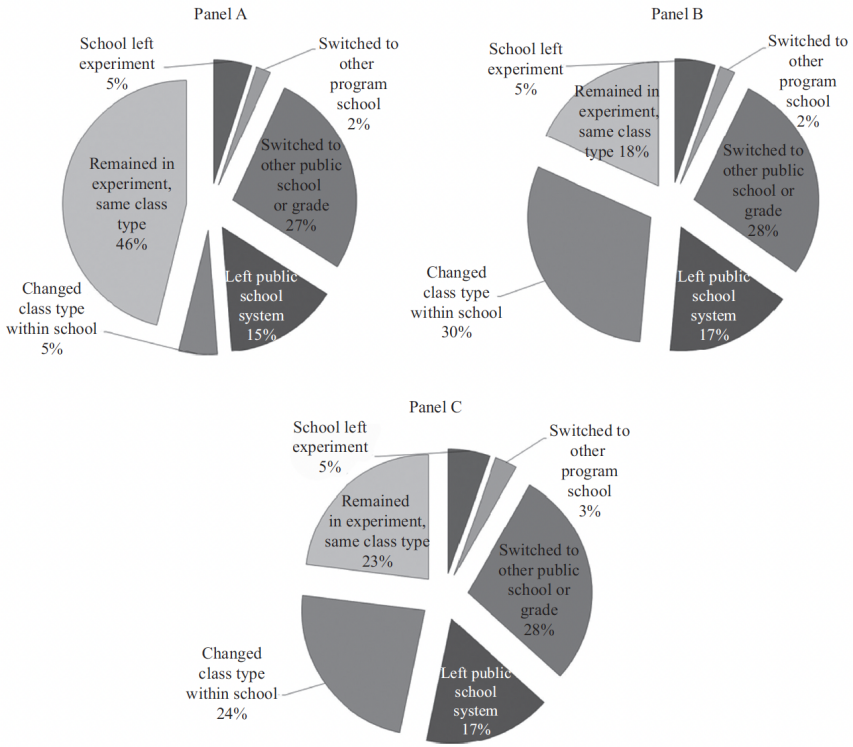


Figure 2: Status in third grade by initial class assignment in Project STAR, kindergarten entry cohort. Panel A: initially in small class, Panel B: initially in regular class, and Panel C: initially in regular class with aide

Notes: Students are counted as having left the public school system if they do not appear in Project STAR in third grade and TCAP scores are not available for them in 1990 (the year in which most students from the cohort were in fourth grade). Additional details in the text.

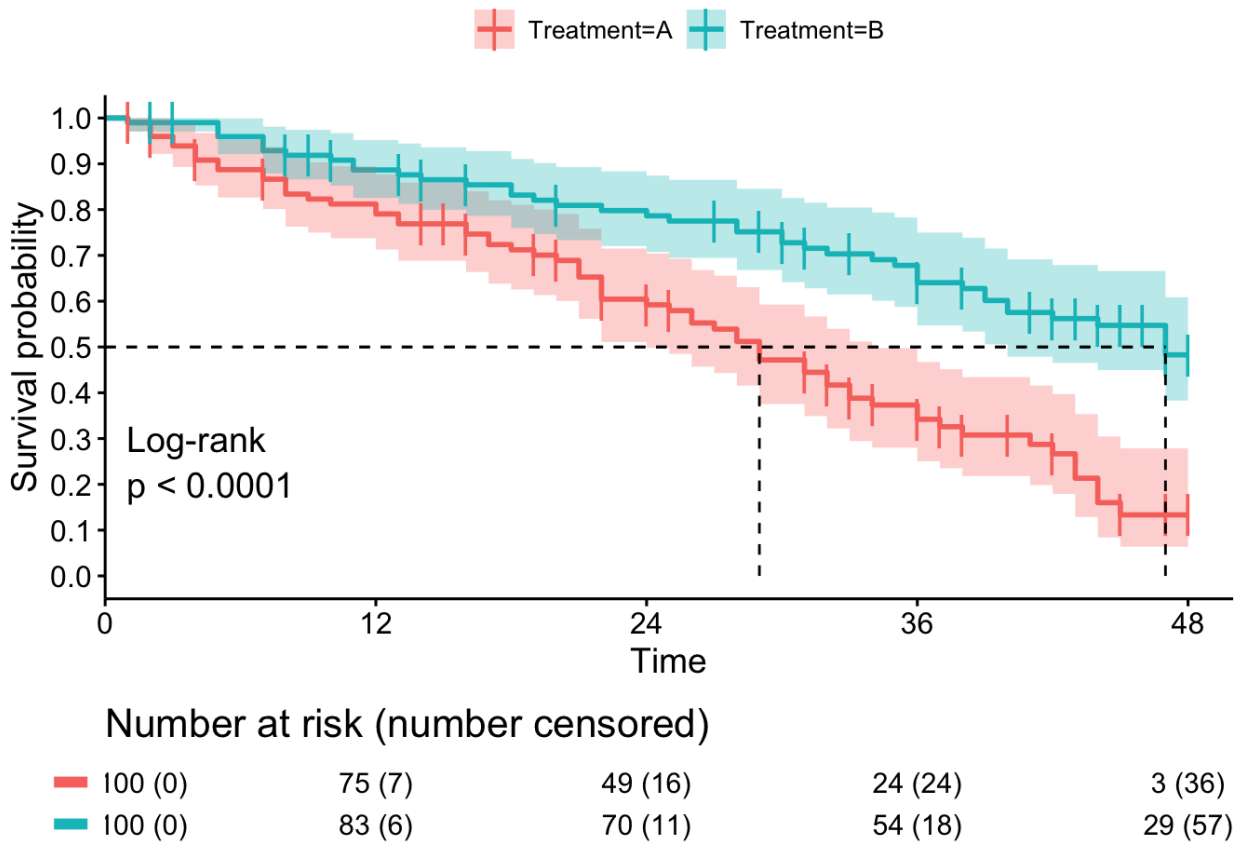
Reproducibility

- Attrition types 1-4 are relatively easy to identify from the data; however, there are issues with type 5
- "Vanishing" may occour through a multitde of processes
  1. Went to a private school
  2. Death (likely small)
  3. Migration to a different state
  4. Not tested (absent, accomodations, etc.)
- Types 1,2,3 are hard to identify and type 4 is perhaps described in the literature

Survival Analysis

1. To quantify the effect of student ability on attrition, I use survival analysis
2. Here, evidence of attrition (falling into category 1, 4, or 5) is considered a "failure"
3. Utilizing $T_{i,g,c,s}$ as the failure time (i.e. number of years a student stayed in the experiment until they left) for each student i in grade g in classroom c at school s
4. To keep the model simple, I assume that attrition is constant across all years Project STAR took place (exponential hazards) and model survival time through an AFT model

Survival Curve



Model

$$log(T_i) = \beta_0 + \beta_1(TS_{i,g,c,s}) + \beta_2(CT_{i,g,c,s}) + \beta_3(X_{i,g,c,s}) + \epsilon_{i,g,c,s} \quad (1)$$

- $TS_{i,g,c,s}$ is the test score for student i in grade g in classroom c in school s
- $CT_{i,g,c,s}$ is a categorical variable for the class type student i was enrolled in upon entry to the experiment (e.g. small, regular-sized, or regular sized with teachers aide)
- $X_{i,g,c,t-1,s}$ is a vector of controls for each student including observable student, teacher, and school data

Interpretabilty

- If $\beta_1 = 1$, then $exp(\beta_1) \approx 2.718$. Holding all other variables constant, an individual with $TS_{i,g,c,s}$ one unit greater than another is expected to stay in the Project STAR experiment 2.718 times longer than the other. In other words, the probability this individual "survived" to time $2.718t$ is the same as the probability that another individual has "survived" to time t
- Perhaps this is better modeled through a logistic regression for exit in year t for $t = 0, 1, 2, ...$(avoids the need to assume constant attrition and allows the relationship between treatment and exit to vary freely with t)

Model

$$TS_{i,g,c,t,s} = \beta_0 + \beta_1(L_{i,g,c,t-1,s}) + \beta_2(A_{i,g,c,t-1,s}) + \beta_3(\overline{CA}_{i,g,c,t-1,s}) + \beta_4(\overline{PA}_{i,g,c,t-1,s}) + \beta_5(X_{i,g,c,t-1,s}) + \beta_6(CT_{i,g,c,t-1,s}) + \beta_7(\overline{PA}_{i,g,c,t-1,s} \times L_{i,g,c,t-1,s}) + \alpha_s + \gamma_{t-1} + \epsilon$$

- $TS_{i,g,c,t,s}$ is the test score for student i in grade g in classroom c in school s at time t
- $L_{i,g,c,t-1,s}$ is the proportion of students in student i's class that left for private schools at time $t-1$
- $A_{i,g,c,t-1,s}$ is student i's ability, measured by their test score in time $t-1$
- $\overline{CA}_{i,g,c,t-1,s}$ is classmate ability for student i. That is, the average test score of the students in student i's class (not including student i) that don't leave the experiment, measured by their test score in time $t-1$
- $\overline{PA}_{i,g,c,t-1,s}$ is the peer ability for student i. That is, the average test score of the students in student i's class that left the experiment in $t-1$
- $X_{i,g,c,t-1,s}$ is a vector of controls for each student including observable student, teacher, and school data
- $CT_{i,g,c,t-1,s}$ is a categorical variable for the class type student i was enrolled in during time $t-1$
- α_s is a school fixed effect, γ_{t-1} is a year fixed effect

Identifying Private School Students

- To identify students who left for private schools, I aim fit a variation of the k-means clustering algorithm (k=?)
- Allows me to separate students into groups that "look" alike thus allowing me to make assumptions about which students went to private school without the data needed to classify these students as private school students
- I use the Mahalanobis distance in clustering to account for covariances between clusters

$$D_m(\vec{v_i}, \vec{v_j}) = \sqrt{\sum_{i=1}^{p} (\vec{v_i} - \vec{v_j})^T \Sigma^{-1} (\vec{v_i} - \vec{v_j})} \ , \ \forall i \neq j$$

Regression for Private School Attrition

$$\begin{aligned} TS_{i,g,c,t,s} = \beta_0 &+ \beta_1(L_{i,g,c,t-1,s}) + \beta_2(A_{i,g,c,t-1,s}) + \beta_3(\overline{CA}_{i,g,c,t-1,s}) \\ &+ \beta_4(\overline{\delta}_{i,g,c,t-1,s}) + \beta_5(\overline{PA}_{i,g,c,t-1,s}) + \beta_6(X_{i,g,c,t-1,s}) \\ &+ \beta_7(CT_{i,g,c,t-1,s}) + \beta_8(\overline{PA}_{i,g,c,t-1,s} \times L_{i,g,c,t-1,s}) + \alpha_s + \gamma_{t-1} + \epsilon \end{aligned}$$

- Not much has changed from previous equation, here $\overline{\delta}_{i,g,c,t-1,s}$ is the cluster peer ability for student i. That is, the average test score of the students in student i's class that left the experiment and the public school system in $t-1$ who, through clustering, I treat as having left for private schools
- Ideally, I would sensitize the k-means to different selections of controls, seeing how the coefficient on $\overline{\delta}_{i,g,c,t-1,s}$ varies based on assumptions I introduce (a variation of what is described in Emily Oster's 2016 paper: *Unobservable Selection and Coefficient Stability: Theory and Evidence*)