# Data

* Additional factors that affected the study design. Beyond the randomization

of students into class types, three operational factors affected the design of the STAR

experiment. First, at the end of the kindergarten year, the STAR consortium decided on

one design modification. There had been no significant differences in the achievement

of regular (R ) classes and teacher-aide (RA) classes in the kindergarten year. Thus,

approximately one-half of R students were randomly assigned to RA classes for the

second year (and beyond), and approximately one-half of RA students were assigned at

random to R classes for the second year (and beyond). No students were purposely

reassigned into or out of small (S) classes. No further modifications of this sort were

made in subsequent years.

* Second, during the summer between grade 1 and grade 2 (summer 1987), a three-day training course was given to 54 second-grade teachers (out of 340) from 15 STAR schools. The training was the same for all 54 teachers, since the assignment to class types had not yet been made. No special attempt was made to prepare teachers to take advantage of a small-class setting. Comparisons of grade-2 achievement scores showed no significant difference between the classes of trained and untrained teachers (see Word et al., 1990, Chapter VI).7 Teachers who participated in the training are flagged in the student data file.
* Third, ordinary student mobility over the years affected the composition and size of STAR classes. Students moving into STAR schools from non-STAR schools during the four-year experiment were assigned at random to one of the class types, with the constraint that small classes could not exceed 17 students.
* Students moving from one STAR school to another were assigned to the same type of class as they had participated in previously (space allowing). Students moving out of a STAR school diminished the class enrollment, occasionally causing the regular classes to become as small as some of the small classes.
* The extent of this “class size drift” is documented in **Achilles (1999)**; its potential impact on statistical results is discussed in **Boyd-Zaharias et al. (1995) and Hedges, Nye, and Konstantopolous (2000).**
* The primary patterns that characterize most STAR students were summarized into a pair of codes in the student data file (Class type composite CMPSTYPE; Duration composite CMPSDURA). These were used in one study to analyze patterns of small-class participation **(Finn, Gerber, Achilles, & Boyd-Zaharias, 2001).**
* Test Scores Administered at the end of each year

# Pie Charts

# Survival Analysis

**Kaplan Meier Plots/Log-Rank Tests (only shown for K-grade entry wave):**

* Dependent variable is “time until attrition” (i.e. Years in star until attrition)
* Function of two things, a dummy for whether or not they left + time until attrition
  + Dummy for attrition calculated based on if years in star != 4 (i.e. did you leave at all)
  + Time in experiment calculated as the sum of the flag variables until first exit (treats students who bobbed in and out as having left the first time they leave)
  + Check differences in Kaplan plots between this and using composite duration
    - Using composite duration variables, we find different results, mostly that being in a small or regular aide
    - Perhaps just use the composite duration variable instead of when they first left? Figure out what’s in this variable or continue to use both and appendix one
* I remove students who leave STAR due to the fact that their school left the experiment, this is because I am focusing on “unforced” attrition (i.e. it was the student’s choice)
* For students that bob in and out, we consider these students right censored. That is, after their first time leaving the experiment, they are considered “left.” ***Here is a summary table of these observations:***

***Income Status***

caption = "Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals."

***Ability***

caption = "Test scores calculated as the sum of kindergarten math, word skills, listening, and reading SAT scaled scores. Percentiles calculated across all students in the kindergarten entry cohort. \n Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals."

***Entry Class Type – Don’t use composite type variables here, switch to a new one***

Caption = “Students may have switched to a different class type during their final year in STAR. \n Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals."

***Entry School Rurality***

Caption = “Students may have switched to a different school during their final year in STAR. \n Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals."

Comparison Plot:

* In the presence of varying class sizes, preferences change based on…
* Show comparison of attrition between in experiment vs. outside experiment

**Survival Model (K-grade entry wave):**

* Role of absent as a control?
* Need to dummy for initial class assignment and then dummy for if they switched classes in year they exited (perhaps improve this by saying what they switched to), but for now just fit the model and see
* 7 observations where kid was in kindergarten, switched to a diff school in 1st, and came back to the same kindergarten school in 2nd. By my estimate, less than .15% of total observations exhibit a similar phenomenon (multiple school switches back to original school)
* Vars of interest:
  + Ability, Free Lunch Status in grade K, entry school urbanicity, switched schools (if this significant, check for final school urbanicity), class assignment in K, switched classes (disaggregate into stayed same, switched from regular to small, switched from small to larger), academic quartiles on entry
* Controls:
  + Gender, Race, years of teacher experience in K, special education in kindergarten, pulled out for special instruction kindergarten, days absent in kindergarten, **kindergarten school grade range, percent students receiving free/reduced lunch in K, percent students bused in K**
  + Note: unable to control for racial composition of school due to the fact that there is too much missing data. All teachers in kindergarten female, thus I don’t control for teacher gender

**Survival Model (1st grade entry wave):**

* Same as before, just diagnostic plots look ass

**Survival Model (comparison students for grade K):**

* How to compare coefficients across groups

# Peer Effects from Attrition

* Filtered out students who switch schools, leave in students that may switch class types. Data on absences not collected in grade 2.
* Bootsrapped SEs?
* Special instruction/education data not collected for grades 2,3

# Clustering (perhaps something on homogeneity in peer quality??)