

Data Appendix

All of the data for this project was supplied by Chicago Public Schools in two parts; first for the violence prediction model and then for evaluation of the YAP mentoring program.

Violence Prediction Model

For the violence prediction model, CPS supplied us with data on all high school students enrolled in the CPS system, totaling over 100,000 students, for the 2007-08, 2008-09, and 2009-10 school years. For 2007-08 and 2008-09, the data included demographics, annual misconducts data, annual attendance data, annual education data, and the list of students who had been victims of a shooting along with the date of the shooting. Using data from September 2007 through April 2009, we calibrated the violence prediction model and developed out of sample predictions for May 2009 through August 2009 and compared the predictions with actual violence data for that period. This period was chosen in order to our model with a prediction model CPS developed in April 2009. We then refit the model using the entire 2008-09 data and used the updated model, which considered separately the most important factors for different race and gender groups, to develop a list for CPS of the highest risk students for the 2009-10 school year. CPS used our list along with similar regression models and a list of students referred by principals (“the principals’ list”), as the basis for referring students to the Youth Advocate Program (YAP) mentoring program that began in second semester of the 2009-10 school year.

YAP Evaluation

The second set of data that we received from CPS was for the approximately 13,000 students who were had highest risk of being victims of a shooting, according to the lists used for referrals to the YAP mentoring program, as well as data for the 361 students who were actually referred to the YAP program. The data for this portion of our project was fairly similar to the data used for the prediction models, except in this case we used more granular data at the semester or monthly level when it was available. The increased granularity was helpful for developing reasonable control and treatment groups for the YAP mentoring program because of the timing of YAP student referrals. The referrals to the YAP program were spread out across both the first and second semesters of 2009-10, however nearly two thirds of the referrals were concentrated into three waves on three different days: October 27, 2009; November 23, 2009; and February 08, 2010.

The treatment groups in our YAP evaluation included only students from those three waves of referrals, since they all received treatment during the majority of the second semester of 2009-10, which began on January 29th. In each of the three waves of referrals, students were referred based on different measures of student risk. The students in the first wave of referrals were referred from a ranking system labeled “prior rank” by CPS. The students in the second wave were referred from the principals’ list only. And students in the third wave were referred from a mix of the principals’ list and CPS’s “revised” ranking system, which was heavily influenced by our list. To account for the different lists used for referrals, we created four separate control groups. The first wave’s control group was simply the next highest-ranked students from the “prior” ranking system who had not been referred to YAP. The control group for the portion of the third wave that was referred from the “revised” ranking system was constructed using the “revised” ranking system in the same way as the control group for the first wave. Control groups for the second wave and the portion of the third wave referred from the principals’ list were constructed using propensity scores. For each student in these waves, a logit regression was run without that student and used to predict the probability of being included on the principals’ list. Then, the non-YAP student with closest predicted probability to the excluded, YAP student was chosen as a control.

Appendix Table 1: Full List of Coefficients for the Violence Prediction Model

	Mean	Full Sample	African American Males	Hispanic Males
Male	0.488 (0.500)	0.003 (0.000)	-	-
African American	0.510 (0.500)	0.001 (0.001)	-	-
Hispanic	0.363 (0.481)	0.000 (0.001)	-	-
Times Shot Previously	0.002 (0.048)	0.008 (0.003)	0.011 (0.007)	-0.012 (0.008)
Serious Misconducts Per Day	0.000 (0.002)	0.347 (0.093)	0.799 (0.231)	-0.327 (0.231)
New to High School	0.250 (0.433)	0.000 (0.000)	-0.001 (0.001)	0.000 (0.001)
English Literacy	0.047 (0.211)	0.000 (0.001)	-0.005 (0.008)	-0.001 (0.001)
Charter School	0.111 (0.314)	-0.001 (0.001)	-0.002 (0.002)	-0.001 (0.002)
Selective School	0.357 (0.479)	0.001 (0.000)	0.000 (0.001)	0.001 (0.001)
Military School	0.018 (0.133)	-0.001 (0.001)	-0.002 (0.004)	0.000 (0.002)
Alternative School	0.032 (0.176)	0.001 (0.001)	0.000 (0.003)	0.012 (0.004)
Special School	0.008 (0.091)	0.000 (0.002)	0.000 (0.006)	-0.001 (0.004)
School Per Capita Shooting History	0.002 (0.005)	0.152 (0.040)	0.119 (0.092)	0.265 (0.117)
Overage	0.100 (0.300)	0.002 (0.001)	0.006 (0.002)	0.002 (0.001)
1 or 2 Credits Behind	0.078 (0.268)	-0.001 (0.001)	-0.001 (0.002)	-0.001 (0.001)
More Than 2 Credits Behind	0.158 (0.365)	0.001 (0.001)	0.002 (0.002)	0.000 (0.001)
Emotional Behavioral Disability	0.027 (0.162)	0.001 (0.001)	0.002 (0.003)	-0.001 (0.003)
Other Learning Disability	0.140 (0.347)	-0.001 (0.001)	-0.004 (0.002)	0.000 (0.001)
Free Lunch Status	0.813 (0.390)	0.000 (0.001)	0.000 (0.002)	0.000 (0.002)
ISAT Math Score	173.113 (124.858)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
ISAT Reading Score	161.060 (116.181)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
% of Days Suspended	0.009 (0.030)	0.030 (0.007)	0.000 (0.018)	0.140 (0.020)
Juvenile Jail	0.015 (0.120)	0.024 (0.001)	0.025 (0.003)	0.014 (0.004)
Adult Detention Center	0.003 (0.059)	0.007 (0.003)	0.009 (0.006)	-0.006 (0.007)
% of Days Absent	0.141 (0.161)	0.012 (0.001)	0.024 (0.003)	0.016 (0.003)

Appendix Table 2: Differences in Fall 2009 Variables For Students Referred to YAP

Outcome Variable	(1)	(2)
Victim of a shooting	0.010 (0.007)	0.011 (0.007)
Dropout	0.019 (0.024)	-0.002 (0.023)
Minor misconducts	0.543** (0.187)	-0.230 (0.173)
Serious misconducts	0.047 (0.053)	-0.066 (0.052)
Courses completed	-0.061 (0.100)	-0.020 (0.098)
Courses failed	0.005 (0.217)	-0.077 (0.195)
GPA	0.002 (0.082)	0.022 (0.072)
Days absent	-3.174 (1.906)	-4.437** (1.646)
Days present	2.282 (2.361)	3.685 (2.085)
Days suspended	0.694 (0.466)	-0.226 (0.447)
Observations	12789	12789
Demographic controls	yes	yes
2008-2009 school year controls	no	yes

Notes: Each entry represents a regression for which the outcome variable was the dependent variable. The given point estimate is the coefficient on a binary variable for being referred to YAP in one of the three early waves, not including students referred from the principal's list. The omitted category is the control group. Standard errors of coefficients are reported in parentheses. * denotes significance at 10%-level; ** denotes significance at 5%-level; *** denotes significance at 1%-level.

Appendix Table 3: Differences in Spring 2009 Variables For Students Referred to YAP

Outcome Variable	(1)	(2)	(3)
Victim of a shooting	-0.000 (0.005)	-0.005 (0.072)	0.005 (0.066)
Dropout	-0.023 (0.035)	-0.142 (0.205)	-0.020 (0.182)
Minor misconducts	0.305* (0.151)	0.122 (1.220)	-0.141 (1.083)
Serious misconducts	0.120** (0.038)	-0.065 (0.251)	-0.035 (0.236)
Courses completed	0.090 (0.242)	-1.126 (1.396)	-0.635 (1.344)
Courses failed	0.052 (0.205)	-1.614 (1.242)	-1.628 (1.340)
GPA	0.024 (0.075)	0.160 (0.380)	0.203 (0.275)
Days absent	1.010 (1.775)	-1.212 (10.413)	4.516 (11.507)
Days present	1.898 (2.398)	11.586 (13.145)	2.497 (12.515)
Days suspended	0.989** (0.381)	0.109 (2.192)	-0.585 (1.953)
Observations	12889	12889	12889
Demographic controls	yes	yes	yes
2008-2009 school year controls	no	yes	yes
Fall 2009 controls	no	no	yes

Notes: Each entry represents a regression for which the outcome variable was the dependent variable. The given point estimate is the coefficient on a binary variable for being referred to YAP in one of the three early waves. The omitted category is the control group. Standard errors of coefficients are reported in parentheses. * denotes significance at 10%-level; ** denotes significance at 5%-level; *** denotes significance at 1%-level.