

A Non-Identification of Dosage Effect

Consider a standard potential outcomes model in which there are three possible “dosages” of SYEP participation: $p = 0, 1, 2$ for 0, 1, or 2 times participating, respectively.¹ Potential outcomes for each state p are given by

$$Y_p = \delta_p + U_p \quad (\text{A-1})$$

where by normalization $E(U_p) = 0$ for all p . The average effect of participating in SYEP once (relative to never participating) is given by $\delta_1 - \delta_0$, and the average effect of participating twice (versus once) is given by $\delta_2 - \delta_1$. We define a “Dosage Effect” as $DE_{21} = (Y_2 - Y_1) - (Y_1 - Y_0)$ and therefore the average dosage effect is $ADE_{21} = (\delta_2 - \delta_1) - (\delta_1 - \delta_0)$. We are interested in testing the whether $ADE_{21} = 0$. Rejecting this restriction with a positive (negative) ADE_{21} would be evidence of supermodularity (submodularity) in SYEP participation: students who participate twice enjoy a larger (smaller) effect of the second participation.

Next, consider attempting to identify ADE_{21} using the data generated by SYEP. For simplicity, we assume full compliance (winning the lottery implies participation). Let $W_p = 1$ denote winning the p^{th} lottery and 0 otherwise, and let $A_p = 1$ denote applying to the p^{th} lottery with 0 otherwise. We can identify at least two local effects with the following estimands:

$$\beta_1 = E(Y|W_1 = 1, A_1 = 1) - E(Y|W_1 = 0, A_1 = 1) \quad (\text{A-2})$$

$$\beta_2 = E(Y|W_2 = 1, W_1 = 1, A_2 = 1, A_1 = 1) - E(Y|W_2 = 0, W_1 = 1, A_2 = 1, A_1 = 1) \quad (\text{A-3})$$

As seen below, β_1 and β_2 correspond to the average treatment effects of the first and second lottery for Groups 1 and 2, respectively, defined in the main text (see Equation 6). Group 1 of first time applicants who have never participated and Group 2 who apply twice after having won and participated once.

Decomposing β_2 , we have

$$\beta_2 = E(Y_2|W_2 = 1, W_1 = 1, A_2 = 1, A_1 = 1) - E(Y_1|W_2 = 0, W_1 = 1, A_2 = 1, A_1 = 1) \quad (\text{A-4})$$

$$= \delta_2 + E(U_2|W_2 = 1, W_1 = 1, A_2 = 1, A_1 = 1) - \{\delta_1 + E(U_1|W_2 = 0, W_1 = 1, A_2 = 1, A_1 = 1)\} \quad (\text{A-5})$$

$$= (\delta_2 - \delta_1) + E(U_2 - U_1|W_1 = 1, A_2 = 1, A_1 = 1) \quad (\text{A-6})$$

where the last equality is given by lottery randomization, which implies that $E(U_p|W_p = 1, \cdot) = E(U_p|W_p = 0, \cdot)$ for all p (and conditioning variables prior to p). Therefore, β_2 does indeed identify the average treatment effect of winning the second lottery for Group 2. $(\delta_2 - \delta_1)$ is the average effect of participating twice relative to once, and $E(U_2 - U_1|W_1 = 1, A_2 = 1, A_1 = 1)$ is the selection effect for Group 2, who apply twice after having won and participated once.

In general, the application decision is likely endogenous, and so

$$E(U_2 - U_1|W_1 = 1, A_2 = 1, A_1 = 1) \neq E(U_2 - U_1) = 0 \quad (\text{A-7})$$

¹Although we do consider three-time applicants in the main analyses, here we restrict our attention to participating at most twice. The point at hand is made no clearer by considering higher “dosages.”

Therefore, comparing the outcomes of the second lottery winners and losers in Group 2 does not identify the average effect of participating twice over once. Further, we see from decomposing β_1 that

$$\beta_1 = (\delta_1 - \delta_0) + E(U_1 - U_0|A_1 = 1) \tag{A-8}$$

The difference in these two estimands identifies

$$\beta_2 - \beta_1 = (\delta_2 - \delta_1) - (\delta_1 - \delta_0) + E(U_2 - U_1|W_1 = 1, A_2 = 1, A_1 = 1) - E(U_1 - U_0|A_1 = 1) \tag{A-9}$$

And it is clear that assuming that the selection terms are equal in magnitude is extremely stringent and unlikely to obtain. We conclude that the average dosage effect is not identified.

B Additional Tables

Table A1: Likelihood of Winning SYEP Lottery by Matching to DOE Data

	Matched	Not Matched	Total
2005	63.7	64.0	63.7
2006	60.8	61.2	60.9
2007	50.5	51.1	50.7
2008	48.5	48.0	48.4

Notes: Applications to vulnerable youth programs, programs based out of the city, or programs with a greater than 99 percent or less than 0 percent selection rate are omitted.

Table A2: Probability of Being Matched to DOE Data (2006 - 2008)

	<i>Dependent variable:</i>
	Matched
Select	0.003 (0.003)
CBO x Year FE?	Y
Grade FE?	Y
Observations	120,817
R ²	0.037

Note: *p<0.1; **p<0.05; ***p<0.01. Heteroskedastic robust standard errors clustered at the lottery-level. Grade is last grade before application which includes 7-12th grade and an additional category for alternative programs. Sample includes all unmatched applications to SYEP in years 2006-2008 and all first-time applicants in those years. 2005 is excluded since we cannot see applications made before 2005 and thus we cannot distinguish first-time applicants from repeat applicants in 2005.

Table A3: Attrition in Year Following Application

Grade	Frac Attrite Winners	Frac Attrite Losers	Frac Attrite All	N Applications
8	2.8	2.9	2.8	20,855
9	4.1	4.5	4.3	50,613
10	2.5	2.8	2.6	42,227
11	4.1	4.0	4.1	23,327
Alt. Program	31.7	30.8	31.2	1,137
Total	3.6	3.9	3.8	138,159

Notes: Attrition is defined as not appearing in NYCDOE administrative data in the year following the SYEP lottery. Main Analyses condition on non-attrition, so number of analyzed applications is mechanically smaller than total presented here. Sample includes all applications for students expected to be in high school following SYEP. Applications are omitted if the student submits multiple applications or in ungraded special education following SYEP. Applications to vulnerable youth programs, programs based out of the city, or programs with a greater than 99 percent or less than 0 percent selection rate are omitted.

Table A4: Impact of Winning Lottery on Attrition (2006 - 2009)

	<i>Dependent variable:</i>					
	8th Grade	9th Grade	10th Grade	11th Grade	Alt Program	All Grades
	(1)	(2)	(3)	(4)	(5)	(6)
Select.prev	0.0003 (0.002)	-0.001 (0.002)	0.001 (0.002)	0.0003 (0.003)	0.026 (0.033)	0.0003 (0.001)
CBO x Year FE?	Y	Y	Y	Y	Y	Y
Cohort FE?	Y	Y	Y	Y	Y	Y
Grade FE?	N	N	N	N	N	Y
Observations	20,855	50,613	42,227	23,327	1,137	138,159
R ²	0.037	0.021	0.032	0.040	0.207	0.038

Note: *p<0.1; **p<0.05; ***p<0.01. Outcome variable is attrition in years 2006-2009. Attrition is defined as not appearing in DOE administrative data in the year following the SYEP lottery. Main analyses condition on non-attrition, so number of analyzed applications is mechanically smaller than total presented here. Sample includes all applications for students expected to be in high school following SYEP. Applications are omitted if the student submits multiple applications or in ungraded special education following SYEP. Applications to vulnerable youth programs, programs based out of the city, or programs with a greater than 99 percent or less than 0 percent selection rate are omitted.

Table A5: Fraction of NYC DOE Students Attempting at Least One Regents Exam 2006 - 2009

Grade	Fraction
8	12.8
9	50.4
10	76.9
11	84.3
12	53.3
Alt. Program	17.7

Notes: NYC DOE students include all students with non-missing grades who appear in administrative data.

Table A6: Comparing ITT Effect Size

Panel A: ITT Estimates							
	Any Attempt	N. Attempts	Any Pass 65	N. Pass 55	N. Pass 65	N. Pass 75	ZScore
Select	0.004	0.021	0.007	0.026	0.018	0.007	0.008
Black	0.011	-0.007	-0.019	-0.067	-0.123	-0.204	-0.149
Free Lunch	-0.036	-0.079	-0.053	-0.105	-0.115	-0.078	-0.074
Zread	0.001	-0.002	0.051	0.092	0.171	0.221	0.245

Panel B: SYEP Effect / Covariate Coefficient							
	Any Attempt	N. Attempts	Any Pass 65	N. Pass 55	N. Pass 65	N. Pass 75	ZScore
Perc Black	0.369	3.004	0.383	0.380	0.148	0.037	0.052
Perc Free Lunch	0.111	0.261	0.137	0.243	0.158	0.096	0.105
Perc Zread	4.772	10.901	0.144	0.278	0.107	0.034	0.032

Notes: For each outcome, the percent of covariate effect (Panel B) is defined as the ITT effect of SYEP divided by the absolute value of the coefficient on a given covariate in the estimated ITT model.

Table A7: Comparing TOT Effect Size

Panel A: ITT Estimates							
	Any Attempt	N. Attempts	Any Pass 65	N. Pass 55	N. Pass 65	N. Pass 75	ZScore
Worked	0.005	0.026	0.009	0.033	0.023	0.010	0.010
Black	0.011	-0.008	-0.019	-0.069	-0.124	-0.205	-0.150
Free Lunch	-0.036	-0.079	-0.053	-0.105	-0.115	-0.078	-0.074
Zread	0.001	-0.002	0.051	0.092	0.171	0.221	0.245

Panel B: SYEP Effect / Covariate Coefficient							
	Any Attempt	N. Attempts	Any Pass 65	N. Pass 55	N. Pass 65	N. Pass 75	ZScore
Perc Black	0.485	3.345	0.484	0.480	0.189	0.047	0.066
Perc Free Lunch	0.143	0.336	0.177	0.313	0.204	0.124	0.132
Perc Zread	6.095	14.351	0.185	0.358	0.137	0.044	0.040

Notes: For each outcome, the percent of covariate effect (Panel B) is defined as the TOT effect of SYEP divided by the absolute value of the coefficient on a given covariate in the estimated TOT model.

Table A8a: Heterogeneous Treatment-on-the-Treated Estimates

	<i>Dependent variable:</i>						
	Any Attempt	N. Attempts	Any Pass 65	N. Pass 55	N. Pass 65	N. Pass 75	ZScore
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Worked	-0.024 (0.055)	-0.121 (0.151)	0.042 (0.055)	-0.111 (0.138)	-0.083 (0.122)	-0.103 (0.093)	0.081 (0.105)
Worked x Female	0.005 (0.006)	0.038** (0.019)	0.005 (0.006)	0.034** (0.017)	0.023 (0.015)	0.011 (0.011)	-0.004 (0.011)
Worked x Black	-0.012 (0.017)	-0.017 (0.051)	-0.019 (0.018)	0.007 (0.049)	0.015 (0.046)	0.015 (0.040)	-0.008 (0.030)
Worked x Asian	-0.002 (0.018)	-0.024 (0.056)	-0.015 (0.019)	-0.003 (0.055)	0.015 (0.052)	0.028 (0.048)	-0.026 (0.033)
Worked x Hispanic	-0.001 (0.017)	0.002 (0.053)	-0.008 (0.018)	0.036 (0.050)	0.049 (0.047)	0.043 (0.041)	0.016 (0.031)
Worked x Free Lunch	0.007 (0.010)	-0.027 (0.032)	-0.003 (0.011)	-0.042 (0.030)	-0.030 (0.027)	0.026 (0.022)	-0.010 (0.018)
Worked x Red Lunch	0.010 (0.013)	-0.036 (0.041)	-0.008 (0.014)	-0.062 (0.039)	-0.041 (0.036)	0.048 (0.030)	-0.017 (0.023)
Worked x LEP	-0.007 (0.022)	-0.006 (0.084)	-0.021 (0.028)	-0.034 (0.076)	-0.060 (0.068)	-0.035 (0.049)	-0.065 (0.048)
Worked x ESL Not LEP	-0.013 (0.030)	0.030 (0.098)	-0.003 (0.030)	0.106 (0.086)	0.055 (0.075)	-0.051 (0.054)	0.069 (0.062)
Worked x Spec Ed	0.016 (0.011)	0.046 (0.033)	0.033*** (0.011)	0.063** (0.028)	0.054** (0.023)	0.012 (0.015)	0.048** (0.024)
Worked x Age	0.002 (0.003)	0.009 (0.008)	-0.001 (0.003)	0.009 (0.008)	0.006 (0.007)	0.003 (0.005)	-0.003 (0.006)
Worked x Zread	0.019*** (0.005)	0.049*** (0.015)	0.010** (0.005)	0.024* (0.013)	0.024* (0.012)	0.025** (0.010)	-0.004 (0.009)
Worked x Zmath	-0.006 (0.005)	0.017 (0.015)	0.001 (0.005)	0.011 (0.013)	0.003 (0.012)	-0.014 (0.010)	-0.003 (0.009)
CBO x Year FE?	Y	Y	Y	Y	Y	Y	Y
Cohort FE?	Y	Y	Y	Y	Y	Y	Y
Grade FE?	Y	Y	Y	Y	Y	Y	Y
Observations	134,366	134,366	134,366	134,366	134,366	134,366	96,200
R ²	0.06	0.064	0.133	0.117	0.169	0.214	0.325

Note: *p<0.1; **p<0.05; ***p<0.01. Heteroskedastic robust standard errors clustered at the student-level. Students in 12th grade, below 8th grade, and in ungraded special education are excluded. Cohort is an indicator for the year of first application to SYEP interacted with the grade of the student when first applied. There are 24 unique cohorts in the sample. Limited English Proficiency (LEP) is determined by score on the Language Assessment Battery exam. Zread and Zmath are 8th grade state test scores, standardized by grade and year of administration. Grade is current grade level in school which includes 8-11th grade and an additional category for alternative specialized programs (for example GED programs)

Table A8b: Joint Test of Treatment Interactions

	Any Attempt	N. Attempts	Any Pass 65	N. Pass 55	N. Pass 65	N. Pass 75	ZScore
F-Stat	1.635	2.537	1.364	1.558	1.392	1.180	1.311
P-Value	0.052	0.001	0.149	0.071	0.135	0.275	0.179

Notes: F-statistics and p-values from test of joint restriction that all treatment-by-covariate coefficients are zero.

Table A9: Expected Benefit Moments - Apply Less Non-Apply

	Any Attempt	N. Attempts	Any Pass 65	N. Pass 55	N. Pass 65	N. Pass 75	ZScore
Avg. EB	-0.522 (0.315)	-0.497 (0.192)	-0.089 (0.671)	-0.061 (0.811)	0.135 (0.678)	-0.056 (0.935)	0.35 (0.239)
P01 EB	-0.119 (0.483)	-0.114 (0.369)	-0.237 (0.267)	-0.376 (0.142)	-0.436* (0.059)	-0.202 (0.346)	-0.498*** (0.009)
P10 EB	0.108 (0.607)	0.112 (0.45)	-0.354 (0.578)	-0.195 (0.57)	-0.377 (0.314)	-0.204 (0.569)	-0.778 (0.102)
P50 EB	-0.845 (0.234)	-0.644 (0.133)	-0.117 (0.581)	-0.061 (0.824)	0.016 (0.961)	-0.249 (0.696)	0.019 (0.95)
P90 EB	-0.111 (0.436)	-0.118 (0.38)	-0.136 (0.249)	-0.134 (0.201)	-0.065 (0.477)	-0.097 (0.41)	-0.034 (0.768)
P99 EB	-0.114 (0.168)	-0.12 (0.126)	-0.137 (0.125)	-0.147 (0.235)	-0.052 (0.495)	-0.07 (0.451)	-0.118 (0.421)

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Estimate presented is moment computed on appliers less that computed on non-appliers, expressed as a percentage of applier moment. Bootstrap p-values in parentheses computed against the null hypothesis of no difference between moment for appliers and non-appliers. For each outcome, expected benefit (EB) is the predicted treatment effect of SYEP given student covariates and the 2SLS estimates of heterogeneous treatment effects by student covariates. Bootstrap standard errors are calculated with 1000 bootstrap iterations, block clustered at the student level. Bootstrap p-values for LATE, mean, and quantiles computed for a two-sided test, and bootstrap p-values for difference in select quantiles computed for a one-sided test.

Table A10a: 90th-10th EB Quantile Differential in Mean Covariates (Applicants)

	Any Attempt	N. Attempts	Any Pass 65	N. Pass 55	N. Pass 65	N. Pass 75	ZScore
Age	0.189*** (0)	0.364*** (0)	-0.645*** (0)	0.685*** (0)	0.276*** (0)	0.337*** (0)	-0.525*** (0)
Asian	0.199*** (0)	0.193*** (0)	0.034*** (0)	-0.004 (0.346)	-0.007* (0.074)	0.035*** (0)	-0.384*** (0)
Black	-0.614*** (0)	-0.406*** (0)	-0.486*** (0)	-0.43*** (0)	-0.463*** (0)	-0.534*** (0)	0.04*** (0)
ESL Not LEP	-0.041*** (0)	0.016*** (0)	-0.007*** (0)	0.135*** (0)	0.083*** (0)	-0.091*** (0)	0.127*** (0)
Female	0.454*** (0)	0.634*** (0)	0.278*** (0)	0.7*** (0)	0.563*** (0)	0.433*** (0)	-0.205*** (0)
Free Lunch	-0.009 (0.144)	-0.265*** (0)	-0.016*** (0.01)	-0.193*** (0)	-0.175*** (0)	0.012* (0.091)	0.02*** (3e - 04)
Hispanic	0.319*** (0)	0.118*** (0)	0.276*** (0)	0.419*** (0)	0.501*** (0)	0.588*** (0)	0.324*** (0)
LEP	-0.029*** (0)	-0.046*** (0)	-0.253*** (0)	-0.155*** (0)	-0.245*** (0)	-0.085*** (0)	-0.27*** (0)
Red Lunch	0.128*** (0)	0.003 (0.452)	-0.122*** (0)	-0.187*** (0)	-0.134*** (0)	0.389*** (0)	-0.109*** (0)
Spec Ed	0.12*** (0)	-0.004 (0.222)	0.561*** (0)	0.257*** (0)	0.27*** (0)	0.044*** (0)	0.781*** (0)
Zmath	0.91*** (0)	2.075*** (0)	1.052*** (0)	1.358*** (0)	1.203*** (0)	0.049*** (1e - 04)	-1.525*** (0)
Zread	2.497*** (0)	2.611*** (0)	1.776*** (0)	1.801*** (0)	1.947*** (0)	1.626*** (0)	-1.686*** (0)

Notes: *p<0.1; **p<0.05; ***p<0.01. Displayed are differences in mean covariates between between top and bottom deciles of expected benefit: $E[X \text{ — EB quantile}=90, \text{outcome}=Y] - E[X \text{ — EB quantile}=10, \text{outcome}=Y]$. P-values for t-test of equality of means displayed. Applicants only.

Table A10b: 90th-10th EB Quantile Differential in Mean Covariates (Non-Appliers)

	Any Attempt	N. Attempts	Any Pass 65	N. Pass 55	N. Pass 65	N. Pass 75	ZScore
Age	0.152*** (0)	0.661*** (0)	-1.318*** (0)	1.38*** (0)	0.524*** (0)	1.014*** (0)	-0.954*** (0)
Asian	0.129*** (0)	0.089*** (0)	-0.142*** (0)	-0.141*** (0)	-0.144*** (0)	-0.03*** (0)	-0.4*** (0)
Black	-0.521*** (0)	-0.38*** (0)	-0.331*** (0)	-0.242*** (0)	-0.173*** (0)	-0.233*** (0)	0.01*** (0)
ESL Not LEP	-0.099*** (0)	0.046*** (0)	-0.026*** (0)	0.378*** (0)	0.208*** (0)	-0.208*** (0)	0.373*** (0)
Female	0.417*** (0)	0.582*** (0)	0.267*** (0)	0.565*** (0)	0.488*** (0)	0.405*** (0)	-0.11*** (0)
Free Lunch	-0.121*** (0)	-0.444*** (0)	-0.298*** (0)	-0.231*** (0)	-0.272*** (0)	0.031*** (0)	-0.05*** (0)
Hispanic	0.111*** (0)	-0.024*** (0)	-0.033*** (0)	0.346*** (0)	0.382*** (0)	0.578*** (0)	0.326*** (0)
LEP	-0.092*** (0)	-0.124*** (0)	-0.604*** (0)	-0.437*** (0)	-0.615*** (0)	-0.173*** (0)	-0.64*** (0)
Red Lunch	0.123*** (0)	-0.009*** (0)	-0.069*** (0)	-0.164*** (0)	-0.095*** (0)	0.373*** (0)	-0.049*** (0)
Spec Ed	0.093*** (0)	0.006*** (0)	0.393*** (0)	0.174*** (0)	0.214*** (0)	0.039*** (0)	0.545*** (0)
Zmath	0.926*** (0)	2.16*** (0)	1.448*** (0)	1.372*** (0)	1.178*** (0)	-0.247*** (0)	-0.314*** (0)
Zread	2.798*** (0)	2.856*** (0)	2.464*** (0)	2.13*** (0)	2.364*** (0)	1.606*** (0)	-0.485*** (0)

Notes: *p<0.1; **p<0.05; ***p<0.01. Displayed are differences in mean covariates between between top and bottom deciles of expected benefit: $E[X \text{ — EB quantile}=90, \text{outcome}=Y] - E[X \text{ — EB quantile}=10, \text{outcome}=Y]$. P-values for t-test of equality of means displayed. Non-appliers only.

Table A11: Joint Test of Zero Treatment Effect on Outcomes 1-6

	Full Sample	Group 1	Group 2	Group 3
F-Stat	2.731	1.322	2.511	1.102
P-Value	0.012	0.243	0.020	0.359

Notes: For each group, statistics generated from F-test of joint restriction that treatment effect is zero for all outcomes other than average Z-score. Average Z-score is omitted as students who attempt no exams have no defined average score. Group 1 consists of all first-time applicants in years 2006-2008. Group 2 is all students who applied for the second time in 2006-2008 and had applied, won, and participated in the prior year. Group 3 is all students who applied for the third time in 2006-2008 and had applied, won, and participated in each of the two years prior. Full Sample is identical to that analyzed in Tables 6-7.