Using Multivariate Matched Sampling That Incorporates the Propensity Score to Establish a Comparison Group

CSE Technical Report No. 596

Denise D. Quigley
Center for the Study of Evaluation
University of California, Los Angeles

April 2003

Center for the Study of Evaluation
National Center for Research on Evaluation,
Standards, and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was funded and supported by the University of California Office of the President (UCOP) for the UCOP Statewide Evaluation (UCOPSE).

USING MULTIVARIATE MATCHED SAMPLING THAT INCORPORATES THE PROPENSITY SCORE TO ESTABLISH A COMPARISON GROUP

Denise D. Quigley

Center for the Study of Evaluation (CSE)
University of California, Los Angeles

Policymakers and educators aim to increase the competitive eligibility of high school students applying to the University of California through the School/University Partnership Program and academic development student programs. Along with this challenge, policymakers and educators want to know whether and to what extent the investment in university Partnerships and academic development programs is making a difference in the choices and lives of students. The University of California Office of the President (UCOP) has funded, as part of its overall evaluation, a research project to identify program effects of outreach work on student achievement using a comparison group design and statistical controls. This report summarizes the literature on matching methods and outlines the strategy used for establishing comparison groups. This methodology is the first step in the evaluation analysis strategy used to measure the systemwide effect(s) of School/University Partnerships and academic development student programs on individual student-level achievement and student-level eligibility.

Overview of Literature on Matched Sampling Methodologies

The most accepted and reliable method to detect treatment effects is a randomized study; however, in educational settings this strategy is virtually impossible to implement. Therefore, when ethical or practical reasons make it impossible to randomly assign units to treatment and control groups, an observational study is conducted. An observational study is an empirical investigation of treatments, policies, or exposures and the effects they cause when assignment of treatments is not controlled. Therefore, when using an observational study, probabilistically equivalent groups must be constructed to estimate effects.

Matched sampling is the common method of selection and adjustment in observational studies (Heckman, 1989; Lalonde, 1986; Manski & Garfinkel, 1992; Rosenbaum, 1989, 1995; Rosenbaum & Rubin, 1983). There are many other methods of estimating effects in observational studies, such as differences–in-differences or

hierarchical linear modeling; however, these other techniques require a census or a large representative sample population to estimate an effect. Because matched sampling is so refined in its selection of controls, effects can be estimated with a relatively small sample population. This is of great benefit when the overall population is small or the costs of collecting data are high.

Matched sampling is a method of selecting units from a large reservoir of potential controls to produce a control group of modest size that is similar to a treated group with respect to the distribution of observed covariates. Multivariate matching methods use several variables to construct probabilistically equivalent groups on observed characteristics. The matching can be done either by matching a whole group or by pairwise matching.

Propensity score matching refers to a class of multivariate matching methods and is the analog of randomization in ideal experiments, despite the fact that propensity score matching is far less complete than randomization and can only balance the distribution of observed covariates (Rubin & Thomas, 2000). Randomization balances on both unobserved and observed covariates, whereas propensity score matching balances only on observed covariates. Despite the fact that propensity score matching is not able to balance unobserved covariates, as randomization can, it is superior to other multivariate matching methods because of its ability to reconcile variation across many observed factors and still establish probabilistically equivalent groups. The use of the propensity score in constructing control groups is widely accepted and considered the optimal method of establishing a comparison group for an observational study (Dehejia & Wahba, 1998; Rosenbaum, 1983, 1995; Rosenbaum & Rubin, 1985).

The limitations of this approach are twofold: the potentially significant amount of unobserved bias, and the potential likelihood of not finding a sufficient match for each treated subject.

Details of the Strategy Used to Establish Comparison Groups

We constructed a control group of elementary, middle and high schools using multivariate matched sampling that incorporated the propensity score using pairwise matching. This method has several steps through which the matching strategy identified for each UC Partner school a similar control school. In total, 72 Partner high schools, 57 Partner middle schools, and 92 Partner elementary schools were included in the initial set of Partner schools for matching.

The first step in the matching process was to identify the potential set of Match schools across the state. Potential Match schools were "regular" public schools that were not currently or had not in the last 3 years been UC Partnership schools and had not participated in the following programs: Advancement via Individual Determination (AVID); Puente; Mathematics, Engineering, Science Achievement (MESA); Early Academic Outreach Programs (EAOP); or the California Subject Matter Projects. To establish the group of regular schools, we first excluded a total of 207 schools in the state that were either alternative (N = 20), community (N = 1), juvenile (N = 1), or charter schools (N = 91), or schools that were not able to be categorized as an elementary, middle or high school (N = 131), or that were missing Academic Performance Index (API) demographic and test score data for 1999/2000 (N = 5). Then we excluded schools that worked with the UC campuses. Across the state, we found the potential set of 275 Match high schools, 559 Match middle schools, and 3,0756 Match elementary schools.

From the group of potential Match schools, we then further identified any school that had similar characteristics to a Partner school on a group of key variables known as "match variables": school quintile on the Stanford Achievement Test, 9th edition (Stanford 9 or SAT9), total school enrollment, percent eligible for free/reduced price meals, percent English language learners (ELL), percent Hispanic, percent African American, percent Asian, and percent White. We considered a school with a mean within one standard deviation of a Partner school's mean to be similar.

The matching process yielded the following results. Of the 92 UC Partner elementary schools, 3 UC Partner elementary schools did not have data and 1 UC Partner elementary school did not have at least one similar school in the potential set of Match schools. These 4 UC Partner elementary schools were not included in the final Elementary Partner Group (N = 88).

Of the 57 UC Partner middle schools, 3 did not have data and 11 did not have at least one similar school in the potential set of Match schools. These two sets of schools—the 11 middle schools without matches and the 43 middle schools with matches—differed only modestly on mean characteristics. The 43 UC Partner schools with a match had higher percentages of African American and Asian students, as well as a smaller number of students entering the school, and a slightly higher mean score on the API in 2000 than the UC Partner schools without matches. The final Middle School Partner Group comprised 43 schools.

Of the 72 UC Partner high schools, 4 did not have data and 35 did not have at least one similar school in the potential set of Match schools. These 35 schools had either very low percentages of White students or no White students, and very high percentages of either Hispanic or African American students—the reason that UC is working with them on academic development. The 4 schools without data and the 35 schools for which we could not find a match were not included in the final High School Partner Group (N = 33). The differences between the UC Partner high schools for which we found matches (N = 33) and did not find matches (N = 35) are examined in the next section, Discussion of Matched Sampling Results.

Finally, from the set of similar schools, we identified the similar school that had the nearest or exact value propensity score to each Partner school's propensity score to become the Match control school. To do this, we first estimated a propensity score (the probability of being a UC Partner school) for every school, including UC Partner schools and all other potential Match schools. The estimated probabilities ranged from 0 to 1. To estimate the probability of being a UC Partner school, we performed a logistic regression using as the dependent variable the dichotomous variable of being a UC Partner (=1) and not being a UC Partner (=0) for all schools in the potential match pool and all Partners. The logistic regression prediction equation included the following independent variables: school quintile on the Stanford 9 (SAT9); total enrollment; total number of certificated teachers; suburban and rural dummy variables with urban as a referent group; percent eligible for free/reduced price meals; percent English language learners; and percent Hispanic, percent African American, and percent Asian, with percent White as a referent group. Note that the prediction logistic regression included more variables than were used in the initial one-standard-deviation matching process. From the logistic regression, each school included in the regression was assigned a predicted value of "being a UC Partner." Then we chose the Match school from the set of potential matches that had the nearest value propensity score to the UC Partner school's score.

Once we chose the Match school with the nearest value propensity score, we established for each Partner school a unique Match school. In the end, the set of 88 UC Partner elementary schools had a set of 88 Non-UC Partner elementary schools that matched across all of the "match variables" and the propensity score. The 43 UC Partner middle schools had a set of 43 Non-UC Partner middle schools that matched across all the variables of interest. The 33 UC Partner high schools had a set of 33 Non-UC Partner high schools that matched across all the variables of interest.

Discussion of Matched Sampling Results

Once the comparison group was constructed using the multivariate matching process that incorporates the propensity score, we compared the two groups in terms of their means and distributions on variables of interest. Table 1 reports the test of equality of the means and distributions for the elementary school Partners and Matches. Table 2 reports the results for the middle school Partners and Matches. Table 3 reports the results for the high school Partners and Matches.

Comparing the elementary UC Partner schools (N = 88) to the elementary Match schools (N = 88; Table 1), we see that the group means of elementary schools did not differ statistically in terms of the probability of being a Partner, percent African American, percent Asian, percent Hispanic, percent White, percent receiving free/reduced price meals, percent English language learners, total enrollment, percent Filipino, the number of students who entered into the school in 1999/2000

Table 1

Test of the Equality of Means and Distributions for Elementary Partners (N = 88) and Elementary Matches (N = 88)

	Equality of distribution (Kolmogorov-Smirnov test)		Equality of means (2-sided)	
Variable	Combined K-S	*p* Value	*t* Statistic	*p* Value
Propensity00	0.0490	1.0000	0.1493	0.8815
Pct_aa	0.0795	0.9220	–0.3207	0.7488
Pct_as	0.0909	0.8200	–0.3070	0.7592
Pct_hi	0.0682	0.9800	0.2181	0.8276
Pct_wh	0.1136	0.5560	0.6361	0.5255
Pct_meal	0.1364	0.3240	–0.7343	0.4638
Pct_ell	0.0909	0.8200	0.0762	0.9394
Enroll	0.1247	0.4440	–0.9366	0.3503
Pct_fi	0.0909	0.8200	–1.3453	0.1803
Mobility	0.0682	0.9800	0.4097	0.6825
ACS_K3	0.1451	0.2580	1.9241	0.0560
ACS_46	0.0895	0.8410	0.7226	0.4709
FULL	0.0682	0.9800	–0.4418	0.6592
EMER	0.1477	0.2370	1.5957	0.1124

Note. Pct = percent; aa = African American; as = Asian; hi = Hispanic; wh = White; meal = free/reduced price meals; ell = English language learners; Enroll = total enrollment; fi = Filipino; Mobility = no. students entering in 1999/2000; ACS = average class size; FULL = full credentialed; EMER = emergency credentialed.

(termed "mobility"), average class size for Grades K-3, average class size for Grades 4-6, total number of full-credentialed teachers, or total number of emergency-credentialed teachers. We assumed unequal variance and tested at the 0.05 significance level. In addition, we tested whether the distributions of these variables were equal using the Kolmogorov-Smirnov (K-S) test. The distributions for the above variables are also statistically similar. Note that if the means differ, then there is evidence that the distributions differ. Any discrepancy between the K-S test and the *t* test of means occurs because the K-S test is less sensitive in identifying differences. The K-S test tests for global differences, whereas the *t* test of means focuses on particular properties.

In addition to the tests of equality of means and distributions, we conducted a logistic regression using the UC Partner elementary schools and the Match elementary schools for a total *N* of 176 to predict "being a UC Partner." If the UC Partner schools and Match schools are similar, then the logistic regression should not be able to predict being a Partner (i.e., distinguish between the Partner and Match schools). We can judge how well the logistic regression was able to predict being a UC Partner by examining the discordant and concordant pairs.[1] We found 44.1% discordant pairs, 43.4% concordant pairs, and 12.4% tied. In a perfectly random assignment of schools, one would expect a roughly equal percentage of discordant and concordant pairs. Also one usually finds somewhat more concordant than discordant pairs, even in a random assignment, because the true coefficients are not known and the estimated coefficients tend to favor higher scores for the Partner group (called capitalizing on chance).

Moreover, this procedure of using a logistic regression predicting "being a UC Partner" allows us to calculate a global measure of discriminability for any two groups by using the logistic regression equation on the two groups and calculating the proportions of concordant and discordant pairs. The coefficient gamma

[1] If there are n1 Partner schools and n2 other schools, n1 + n2 = n. Each school has two values: (a) 1 or 0 indicating Partner or other, and (b) the probability of being a Partner based on the logistic regression. If all possible pairs were formed there would be n(n-1)/2 pairs. Of the pairs, many would be Partner-to-Partner and other-to-other; thus, pairs of both 0s or both 1s. Otherwise, each of the n1 Partner schools is paired with each of the n2 other schools, so there are n1*n2 pairs. In this pairing of Partner to other schools, if the Partner school has a higher predicted probability of being a Partner than the other school, then the pair is concordant (i.e., the two variable values go in the same direction). On the other hand, if the Partner school in the pair has a lower predicted probability of being a Partner school than the other school, the pair is discordant. If the two schools have the same predicted probability of being a Partner, the pair is tied. If schools are randomly divided into two groups (Partner and other) and the true best regression equation is used, we expect an equal number of concordant and discordant pairs.

[coefficient gamma = (proportion concordant – proportion discordant) / (proportion concordant + proportion discordant)] is a frequently used correlation-like measure of the discriminability of the groups, varying from +1 (perfect prediction) through 0 (random prediction) through –1 (perfect negative prediction). The group of 176 elementary UC Partner and Match schools yielded a coefficient gamma of –0.008 (essentially zero), indicating that the regression was not able to predict the difference between these UC Partner schools and Match schools.

Thus, the Match elementary schools were not distinguishable from the Partner elementary schools according to the composite of the variables used in the logistic regression. All indications suggested that the groups were therefore similar across the observed variables. Therefore, we found very good, indistinguishable elementary school matches.

Comparing the middle school UC Partner schools (N = 43) to the middle school Match schools (N = 43; Table 2) in the same way, we also see that the group means of the middle schools did not differ statistically in terms of the probability of being a Partner, percent African American, percent Asian, percent Hispanic, percent White, percent receiving free / reduced price meals, percent English language learners, total enrollment, percent Filipino, mobility, or total number of emergency-credentialed teachers. They differed only in terms of the total number of full-credentialed teachers. The K-S test indicated that the distributions of these variables were equal and therefore statistically similar.

In addition to the tests of equality of means and distributions, we conducted a logistic regression using the UC Partner middle schools and the Match middle schools for a total N of 86. We predicted “being a UC Partner” and examined the discordant and concordant pairs. We found 34.4% discordant pairs, 65.1% concordant pairs, and 0.5% tied. The group of 86 UC Partner middle schools and Match schools yielded a coefficient gamma of 0.308, indicating that the regression was able to predict the difference between the UC Partner middle schools and the Match middle schools. Thus, the Match middle schools were slightly distinguishable from the UC Partner middle schools according to the composite of the variables used in the logistic regression. However, since the means and distributions of the two groups were statistically similar across the observed variables, the two groups matched very well and could be used as comparison groups. Therefore, we found slightly distinguishable middle school matches.

Table 2

Test of the Equality of Means and Distributions for Middle School Partners (N = 43) and Middle School Matches (N = 43)

	Equality of distribution (Kolmogorov-Smirnov test)		Equality of means (2-sided)	
Variable	Combined K-S	*p* Value	*t* Statistic	*p* Value
Propensity98	0.1229	0.9050	–0.6520	0.5162
Pct_aa	0.0930	0.9920	–0.2592	0.7961
Pct_as	0.1395	0.7970	–0.6814	0.4975
Pct_hi	0.0698	1.0000	0.2015	0.8408
Pct_wh	0.1860	0.4460	0.4741	0.6366
Pct_meal	0.1163	0.9930	0.6767	0.5005
Pct_ell	0.0930	0.9920	0.0851	0.9324
Enroll	0.1528	0.7040	0.4886	0.6264
Pct_fi	0.1628	0.6190	0.1982	0.8434
Mobility	0.1628	0.6190	1.5092	0.1350
FULL	0.2791	0.7070	**–2.0482**	**0.0437****
EMER	0.2093	0.3030	1.6294	0.1070

Note. Pct = percent; aa = African American; as = Asian; hi = Hispanic; wh = White; meal = free / reduced price meals; ell = English language learners; Enroll = total enrollment; fi = Filipino; Mobility = no. students entering in 1999 / 2000; FULL = full credentialed; EMER = emergency credentialed.

Finally comparing the high school UC Partner schools (N = 33) to the high school Match schools (N = 33; Table 3), we also see that the group means of the high schools did not differ statistically in terms of the probability of being a Partner, percent African American, percent Asian, percent Hispanic, percent White, percent receiving free / reduced price meals, percent English language learners, total enrollment, percent Filipino, mobility, average class size in the core curriculum high school courses, total number of full-credentialed teachers, or total number of emergency-credentialed teachers. We assumed unequal variance and tested at the 0.05 significance level. The K-S test indicated that the distributions of these variables were equal and therefore statistically similar.

In addition, we conducted the logistic regression using the UC Partner high schools and the Match high schools for a total N of 66. We predicted "being a UC Partner" and examined the discordant and concordant pairs. We found 34.8% discordant pairs, 63.4% concordant pairs, and 1.8% tied. The group of 66 UC Partner high schools and Match schools yielded a coefficient gamma of 0.291, indicating that

Table 3

Test of the Equality of Means and Distributions for High School Partners (N = 33) and High School Matches (N = 33)

	Equality of distribution (Kolmogorov-Smirnov test)		Equality of means (2-sided)	
Variable	Combined K-S	*p* Value	*t* Statistic	*p* Value
Propensity98	0.2764	0.2270	–0.1621	0.1107
Pct_aa	0.0909	0.9990	–0.3815	0.7041
Pct_as	0.1818	0.6460	–0.1854	0.8535
Pct_hi	0.1515	0.8430	–0.5163	0.6075
Pct_wh	0.1515	0.8430	1.0438	0.3005
Pct_meal	0.1818	0.6460	–0.1184	0.9061
Pct_ell	0.3030	0.0970	–1.8152	0.0742
Enroll	0.1799	0.6690	–0.0789	0.9374
Pct_fi	0.2424	0.2870	1.2533	0.2147
Mobility	0.1515	0.8430	–0.3981	0.6919
ACS_core	0.0833	1.0000	0.2743	0.7847
FULL	0.2727	0.1720	–1.9030	0.0606
EMER	0.2424	0.2870	1.4776	0.1444

Note. Pct = percent; aa = African American; as = Asian; hi = Hispanic; wh = White; meal = free / reduced price meals; ell = English language learners; Enroll = total enrollment; fi = Filipino; Mobility = no. students entering in 1999 / 2000; ACS = average class size; FULL = full credentialed; EMER = emergency credentialed.

the regression was able to predict the difference between the UC Partner high schools and the Match high schools. Thus, the Match high schools were slightly distinguishable from the UC Partner high schools according to the composite of the variables used in the logistic regression. However, the means and distributions of the two groups were statistically similar across the observed variables, indicating that the two groups matched very well and could be used as comparison groups. Therefore, we found slightly distinguishable high school matches.

In summary, for the schools for which we found potential matches, we were able to find a unique match. For the UC Partner elementary schools, this resulted in a full match. For the UC Partner middle schools, the match was a partial match that yielded, however, a representative group. Table 4 reports the means and *t*-test results for the UC Partner middle schools with a match and those without a match. Finally, for the UC Partner high schools, we were not able to find a potential or unique match for 35 of the 68 UC Partner high schools. These 35 UC Partner high schools also differed systematically from the 33 UC Partner high schools for which

Table 4

Means and Test of the Equality of Means for Middle School Partners With a Match and Without a Match

Variable	Means		Equality of means (2-sided)	
	Partners with match (N = 43)	Partners without match (N = 11)	t Statistic	p Value
Propensity98	0.20	0.28	1.78	0.081
Pct_aa	8.62	29.30	**4.72**	**<0.001**
Pct_as	6.70	15.60	**2.55**	**0.014**
Pct_hi	61.26	40.90	–2.32	0.240
Pct_wh	17.58	9.10	–1.85	0.070
Pct_meal	59.44	73.00	1.99	0.052
Pct_ell	32.26	39.30	1.01	0.316
Enroll	929.05	1037.50	0.90	0.370
Pct_fi	2.93	1.70	–0.68	0.501
Mobility	15.37	24.40	**2.32**	**0.024**
ACS_46	28.50	28.25	–0.19	0.857
ACS_core	27.93	26.40	–1.66	0.104
FULL	83.72	85.30	0.47	0.637
EMER	15.42	14.80	–0.19	0.847
API score 2000	567.19	511.90	**–2.38**	**0.020**

Note. Pct = percent; aa = African American; as = Asian; hi = Hispanic; wh = White; meal = free / reduced price meals; ell = English language learners; Enroll = total enrollment; fi = Filipino; Mobility = no. students entering in 1999 / 2000; ACS = average class size; FULL = full credentialed; EMER = emergency credentialed; API = Academic Performance Index. Boldface type indicates that the means are statistically different at the 0.05 level.

we did find matches. Note that, of the total sample of 72 UC Partner high schools, 4 schools did not have data. Table 5 reports the means and *t*-test results for variables of interest for those UC Partner high schools with a match and without a match.

The UC Partner high schools for which we could not find a match were statistically different from the UC Partner high schools for which we could find a match. Some of the UC Partner high schools were unique, and the schools for which we could not find a match had a higher percentage of African American, Asian, and Hispanic students and a lower percentage of White students; a higher percentage of students receiving free / reduced price meals; a higher percentage of English language learners; greater mobility; fewer full-credentialed teachers and more emergency credentialed teachers; and a lower API based on the Standardized Testing and Reporting (STAR) examinations in 2000.

Table 5

Means and Test of the Equality of Means for High School Partners With a Match and Without a Match

Variable	Means		Equality of means (2-sided)	
	Partners with match (N = 33)	Partners without match (N = 35)	t Statistic	p Value
Propensity98	0.172	0.462	**6.3086**	**0.0000**
Pct_aa	9.515	21.371	**2.8033**	**0.0066**
Pct_as	9.666	6.942	–1.0789	0.2846
Pct_hi	48.878	61.428	**2.0837**	**0.0411**
Pct_wh	25.484	5.685	**–6.7638**	**0.0000**
Pct_meal	37.909	60.085	**4.2517**	**0.0001**
Pct_ell	20.727	33.714	**3.8455**	**0.0003**
Enroll	1473.340	2033.800	**3.2514**	**0.0018**
Pct_fi	2.575	2.000	–0.5977	0.5521
Mobility	15.212	16.657	0.4030	0.6882
ACS_core	27.272	27.735	0.7169	0.4760
FULL	85.848	78.771	**–3.2455**	**0.0018**
EMER	14.272	18.085	**2.1303**	**0.0369**
API score 2000	560.600	482.900	**–5.1223**	**0.0000**

Note. Pct = percent; aa = African American; as = Asian; hi = Hispanic; wh = White; meal = free / reduced price meals; ell = English language learners; Enroll = total enrollment; fi = Filipino; Mobility = no. students entering in 1999 / 2000; ACS = average class size; FULL = full credentialed; EMER = emergency credentialed; API = Academic Performance Index. Boldface type indicates that the means are statistically different at the 0.05 level.

Conclusion

Based on this methodology and the results, we were able to find a set of elementary schools in the state that were statistically similar to the UC Partner elementary schools and a set of UC Partner middle schools in the state that were practically statistically similar to the UC Partner middle schools. This enabled a simple comparison of the UC Partner and Match elementary schools in terms of the means and distributions for outcome variables and achievement measures of interest. Because of the slight differences we found for the middle and high schools, we can compare these two groups of UC Partner schools and their Match schools using analysis of covariance on key variables to remove the linear effects of those remaining differences. All three groups of Partner and Match schools—elementary, middle and high schools—can also be compared using pair differences adjusted by the covariance. It must also be remembered that for the UC Partner high schools, we

were able to find statistically similar schools for only about half (33 out of 68) of that group. We could not find matches for UC Partner high schools with highly disadvantaged populations, primarily because UC is working with all of these schools either through AVID, EAOP, MESA, Puente, UC Partnerships, or the California Subject Matter Projects. Therefore, high school comparisons can be made across the 33 UC Partner high schools and their 33 Match high schools, but these comparisons do not reflect the changes in the entire group of UC Partner high schools.

References

Dehejhia, R. H., & Wahba, S. (1998). *Propensity score matching methods for non-experimental causal studies* (Working Paper No. 6829). Cambridge, MA: National Bureau of Economic Research.

Heckman, J. J. (1989). Casual inference and nonrandom samples. *Journal of Educational Statistics, 14,* 159-168.

Lalonde, R. (1986). Evaluating the econometric evaluations of training programs. *American Economic Review, 76,* 604-620.

Manski, C. F., & Garfinkel, I. (1992). Introduction. In C. Manski & I. Garfinkel (Eds.), *Evaluating welfare and training programs* (pp. 1-22). Cambridge, MA: Harvard University Press.

Rosenbaum, P. R. (1989). Optimal matching in observational studies. *Journal of the American Statistical Association, 84,* 1024-1032.

Rosenbaum, P. R. (1995). *Observational studies.* New York: Springer-Verlag.

Rosenbaum, P. R., & Rubin D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41-55.

Rosenbaum, P. R., & Rubin D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39,* 33-38.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of American Statistical Association, 95,* 573-585.

Appendix

Example of Multivariate Matched Sampling Using the Propensity Score

This appendix guides the reader through an example of how the matching process works given a UC Partner school with the characteristics detailed in Table A1. In Table A1, column 2 shows the Partner's values for the match variables. Column 3 shows the population standard deviation for a given variable, and column 4 shows the range that potential match schools must fall within to be considered a match. The lower boundary of the range for a given variable is calculated by subtracting the Partner population standard deviation from the individual Partner's value for that variable. The upper boundary of the range is calculated by adding the Partner population standard deviation to the individual Partner's value for that variable. For example, this Partner school's enrollment is 1,456 students. By subtracting and adding 961.824 from and to 1,456, the range of 494.18 to 2,417.82 is obtained. This selection can be done for all variables simultaneously using a built-in feature in Microsoft Excel.[2]

Table A1

An Example of the Matching Process

Variable	Partner school	Population *SD*	Range[a]
Propensity98	0.887	0.170	Not applicable
Enrollment	1,456	961.824	494.18 to 2,417.82
Pct_ell	0.199	0.150	0.049 to 0.349
Pct_F_R_meals	0.315	0.198	0.120 to 0.513
Pct_hispanic	0.090	0.263	0.000 to 0.353
Pct _africanam	0.431	0.194	0.236 to 0.625
Pct_asian	0.223	0.150	0.072 to 0.373
Pct_white	0.224	0.156	0.068 to 0.380
SAT quint	3	0.980	Not applicable

Note. Pct = percent; aa = African American; as = Asian; hi = Hispanic; wh = White; F_R_meals = free / reduced price meals; ell = English language learners; Enrollment = total enrollment; fi = Filipino; ACS = average class size; FULL = full credentialed; EMER = emergency credentialed; SAT quint = Stanford Achievement Test quintile.

[a]If the lower bound of the range is a negative number it is set to zero because it is impossible for these variables to have negative values.

[2] The advanced filter in Microsoft Excel was used in the matching process.

Table A2 displays the results of the first step of the matching process. Notice that the two possible matches have values within the selected ranges for all of the match variables.

The next step is to identify the schools that match on SAT quintile. From Table A1, the SAT quintile of the Partner school is 3. Therefore, for the two possible matches to be considered potential Match schools, they must have an SAT quintile of 3, 4, or 5. They both have an SAT quintile of 3, 4, or 5; therefore, these schools are considered potential Match schools. Since the Partner school in this example has at least one potential Match school, this part of the matching process is complete for this school.

Finally, the propensity score for the Partner school is compared to the propensity score of each of its potential match schools. The potential Match school with the nearest (or exact) propensity score value is selected as the match for the Partner school. In our example, we compare the Partner's propensity score of 0.887 to the propensity scores values of 0.761 and 0.823. The propensity score of 0.823 for potential Match 2 is nearer to 0.887; therefore, potential Match 2 is chosen as the final match for the Partner school. Once a match is selected for a Partner school, this school is excluded from being a match to any other of the Partner schools.

Table A2

An Example of the Matching Process—Results of the First Step

Variable	Range[a]	Potential Match 1	Potential Match 2
Propensity98	Not applicable	0.761	0.823
Enrollment	494.18 to 2,417.82	1,323	1,141
Pct_ell	0.049 to 0.349	.145	.162
Pct_F_R_meals	0.120 to 0.513	.308	.371
Pct_hispanic	0.000 to 0.353	0.314	0.112
Pct _africanam	0.236 to 0.625	0.256	0.288
Pct_asian	0.072 to 0.373	0.093	0.171
Pct_white	0.068 to 0.380	0.261	0.239
SAT quint	Not applicable	5	4

Note. Pct = percent; aa = African American; as = Asian; hi = Hispanic; wh = White; F_R_meals = free / reduced price meals; ell = English language learners; Enrollment = total enrollment; fi = Filipino; ACS = average class size; FULL = full credentialed; EMER = emergency credentialed; SAT quint = Stanford Achievement Test quintile.

[a]If the lower bound of the range is a negative number it is set to zero because it is impossible for these variables to have negative values.