Training and Technical Assistance
Webinar Series

EVALUATING CRIME AND JUSTICE POLICIES AND PROGRAMS: USING MATCHING TO APPROXIMATE RANDOMIZED EXPERIMENTS

June 25, 2014

Justice Research and Statistics Association
720 7th Street, NW, Third Floor
Washington, DC 20001
www.jrsa.org

Evaluating Crime and Justice Policies and Programs:

Using Matching to Approximate Randomized Experiments

JRSA Webinar
June, 25, 2014
2:00 – 3:30pm

Bill Bales
College of Criminology and Criminal Justice
Florida State University

INTRODUCTION

- The purpose of the vast majority of research in the field of criminal justice and criminology is to assess the impact of some policy or practice on an outcome of interest.
- Therefore, we strive to compare the outcomes of two groups of individuals who are the same except for one group has received the intervention and the other has not.
- The “Gold Standard” of Random Assignment of subjects to a control group, i.e., those who were not exposed to an intervention, or an experimental group, i.e., those who are exposed to the intervention, is the only method to know whether an intervention has a causal effect on an intended outcome.

- However, there are several impediments to conducting random assignment in the criminal justice field.
- The ethical issue of denying a treatment or an intervention intended to improve individuals under the control of a law enforcement, judicial, or correctional system makes random assignment difficult to implement.
- Designing and implementing the process of informed consent and random assignment requires significant resources and staff time that agencies typically are not willing or not able to devote.
- One exception is the Florida Department of Corrections random assignment study to evaluate the effectiveness of Substance Abuse Treatment on post-prison recidivism.

- What are the typical alternatives to Random Assignment?

- For many decades, the vast majority of research in the field of criminal justice and criminology has compared outcomes across non-equivalent comparison groups, control and experimental, and used some form of regression method as part of the data analysis in an attempt to control for group differences.

- These can take the form of Ordinary Least Squared Regression, Logistic Regression, Survival Analysis, etc.

- The purpose of these modeling methods is to “control” for variables that may influence an outcome of interest to measure the “unique” effect of the key independent variable of interest on the outcome.
- For example, one is trying to determine if drug treatment in prison has a causal effect on the likelihood of post-prison recidivism.
- To accomplish this goal, a host of other factors known to influence recidivism, such as gender, age at release, current offense, prior record, etc. are included in a multivariate statistical model.
- Conclusions are then derived from the model that, after controlling for other factors, drug treatment has a unique effect on reducing or increasing recidivism, or has no effect.

What are the disadvantages of Regression Based Models?

- The regression coefficient representing the effect of your intervention variable on the outcome only represents an “average” treatment effect among the target sample/population rather than the effects of one or more key independent variable(s) across two equivalent groups (control and experimental).
- An important component of policy research is to convince non-researchers that the method used to determine the effect of a policy or intervention on an outcome is sound and the results obtained represent a true and unique effect of the intervention.
- However, explaining how a multi-variate model makes an intervention and non-intervention group equivalent is almost impossible.
- The effect size of your intervention measured in the form of a regression coefficient, odds ratio, etc. may be difficult for non-statisticians to understand, which can result in your findings not being fully understood by practitioners and policy makers.

Advantages of Matching Methods to Derive Equivalent Control and Experimental Groups

- An advantage of alternative estimation techniques such as matching methods is their ability to compare similarly or near similarly situated individuals across conditions, thereby attempting to simulate the effects of an intervention if cases were randomly assigned to a treatment group or a control (non-treatment) group.

- While matching approaches are only as promising as the number of covariates used in the matching and balancing procedure, balance or comparability has advantages over holding constant the effects of extraneous variables through a regression analysis.

- One important issue when using matching procedures is the bias-variance trade-off. Namely, as two different groups are matched and balanced based on an array of characteristics, the matched cases are more homogenous. However, unmatched cases are eliminated from the analysis as they “appear different”. Whether this exclusion reduces the generalizability of the findings is an issue.

Three Types of Matching Methods

- **Propensity Score Matching (PSM)**
- **Precision (Exact) Matching**
- **Coarsened Exact Matching (CEM)**

PROPENSITY SCORE MATCHING(PSM)

- Propensity Score Matching (PSM), first developed by Rosenbaum and Rubin (1983), enables one to compare outcomes across treatment and comparison groups that are as close to equal as possible.
- The basic goal of matching is to create a comparison group as close to equal as possible to the treatment group with respect to observable covariates; thus, the method typically performs best when a large pool of possible comparison subjects exists from which to select.
- Essentially, you are trying to replicate the conditions of an experiment such that the treatment variable can be treated as though it occurred at random and that the individuals in your sample are homogenous on all the measured factors except for the treatment variable.
- Although the use of propensity score matching to determine the effect of interventions when using observation data is more established in such disciplines as medicine and economics (see, e.g., Austin, 2008), it increasingly has been applied in criminology (e.g., Apel et. al., 2007; King, Massoglia, and MacMillan, 2007; Massoglia, 2008; Mears and Bales, 2009).

The Propensity Score in PSM

- The propensity score is the probability of being in the treatment group given that each case has certain characteristics based on observable variables.
- Typically, the propensity score (P-Score) is generated through a logistic regression model. P-Scores range from 0 to 1.
- A pair of treated and control cases sharing the same or similar propensity score can be viewed as comparable even though they may differ on values of specific covariates.
- These scores adjust for, but do not totally solve the problem of selection bias into the treatment group. However, through matching on many observed variables, we can minimize selection bias into an experimental group.
- Through matching individuals in the control and experimental groups based on their propensity scores, i.e., probability of being treated, we derive balance, or equivalence, in the two groups.

Balancing in PSM

- The goal is to create homogenous control and experimental groups based on your treatment or intervention as if the assignment of the groups was random.
- You are balancing treated and untreated cases on their characteristics, which rules them out as confounders and allows for a comparison of the treated and untreated cases with similar propensities to receive the treatment.
- If balance is achieved, then for a given propensity score, exposure to treatment is random and therefore treated and control units should be on average observationally identical.
- In other words, observations with the same propensity score must have the same distribution of observable characteristics independent of their treatment status.
- Therefore, treated and control cases should be, on average, observationally identical.

Types of Propensity Score Matching:

- **One-to-One Nearest Neighbor** – This approach identifies the propensity score of each treatment case and then chooses a control case whose score is closest. You can set a caliper based on the level of precision you want in terms of matching treatment and control cases based on the P-Score. Additionally, you can use replacement or non-replacement of matched control cases to each treatment case.

- **Kernal Matching** – This method is similar to one-to-one nearest neighbor, however, the closer the p-scores values are between experimental and control cases, the more weight they are given in the analysis.

- **One-to-One Pair Matching** – When the propensity scores for the control case have to be identical to the score for an experimental case. This can involve retaining only one matched control cases or a variable number of controls.

Example of Research Using Propensity Score Matching, Mears and Bales, 2009:

- This research examined the effect of supermax incarceration on recidivism. The issues addressed included:
- 1. Whether inmates exposed to supermax incarceration were more likely to recidivate.
- 2. Whether supermax incarceration results in a greater likelihood of violent recidivism.
- 3. Whether longer lengths of stay in supermax incarceration increases recidivism.
- 4. Whether stays in supermax confinement more recent to prison release results in higher recidivism rates.

Table 1. Supermax versus Non-Supermax Inmates: Comparison of Covariate Means in Unmatched and Matched Samples

	Supermax Inmates	Unmatched Sample: Non-Supermax Inmates	Matched Sample: Non-Supermax Inmates
Age at Release	28.50 (.19)	32.25 (.04)	28.36 (.20)
Black	0.75 (.01)	0.59 (.00)	0.75 (.01)
Hispanic	0.05 (.01)	0.06 (.00)	0.04 (.01)
Current Offense Violent	0.54 (.01)	0.39 (.00)	0.55 (.01)
Current Offense Property	0.27 (.01)	0.30 (.00)	0.27 (.01)
Current Offense Drug	0.13 (.01)	0.25 (.00)	0.13 (.01)
Prior Recidivism	0.95 (.03)	0.98 (.00)	0.92 (.03)
Prior Violent Crimes	2.15 (.04)	1.64 (.01)	2.16 (.04)
Prior Escape Convictions	0.12 (.01)	0.07 (.00)	0.13 (.01)
Years in Prison	5.93 (.07)	3.53 (.01)	5.96 (.07)
Violent Discipline Reports	2.69 (.02)	0.62 (.00)	2.69 (.02)
Defiance Discipline Reports	7.44 (.09)	1.34 (.01)	7.47 (.09)
Minimum 85% Law	0.15 (.01)	0.39 (.00)	1.55 (.01)
N	1,241	57,245	1,241

Table 2. Average Effect of Supermax Exposure on Recidivism Outcomes at Three Years Using One-to-One Nearest-Neighbor Matching Without Replacement

	Supermax	Non-Supermax	Difference	*t* Value
Any Recidivism				
Unmatched	58.8%	46.6%	12.2%	
Matched	58.8%	57.6%	1.4%	.69
Violent Recidivism				
Unmatched	24.2%	10.9%	13.3%	
Matched	24.2%	20.5%	3.7%	2.22*
Property Recidivism				
Unmatched	21.5%	17.9%	3.6%	
Matched	21.5%	20.9%	0.6%	0.39
Drug Recidivism				
Unmatched	22.0%	20.2%	1.8%	
Matched	21.9%	22.4%	-0.5%	-0.29

* p < .05 ** p < .01 *** p < .001 (two-tailed test)

PRECISION (EXACT) MATCHING

- Precision or exact matching is one method available for researchers who are trying to improve causal inference, but who cannot randomly assign people to receive an intervention due to ethical and practical constraints.
- Precision (or exact) matching attempts to match the values of all key covariates among the cases in the control and experimental groups.
- Also known as "variable-by-variable" matching (Nagin et al., 2009, p. 145), this method selects cases within the treatment and control group where the values of each matching variable are identical. Those cases that do not match precisely are pruned from the analysis.
- According to Nagin et al. (2009:138), 'exact matching' is the "surest way of accounting for a variable that may somehow be biasing results.", i.e., the next best thing to random assignment.
- The availability of matches on multiple dimensions rapidly moves to the null set as more dimensions of matching are added.

Precision (Exact) Matching (Cont.)

- In practice, the capacity to control for confounders via "by-variable" matching is limited by the "tyranny of dimensionality".
- As more variables are added to precision matching process, fewer matches can be identified.
- This "tyranny of dimensionality" is further exacerbated when matching on continuous variables such as age, prior criminal record, etc..
- Matching on continuous variables such as age, income, etc. typically requires converting the variable into range categories.
- More fine-grained ranges provide for better control but also make it more difficult to identify suitable matches.

Conducting Precision (Exact) Matching

There are two approaches to conducting precision matching. The first to be discussed is through matching across equivalent cases in the control and experimental group using a statistical/database package such as SPSS, SAS, Stata, etc. The second, which is much easier, is using Coarsened Exact Matching (CEM), which will be discussed later.

- 1. Identify variables you want to control for through the matching process and make them character variables and then concatenate multiple variables into one variable.
- 2. Create a dataset that contains your control cases and a dataset that contains cases in the experimental group.
- 3. Match cases across these two files that have the same values on all of the covariates using the concatenated variable.
- 4. Compare the outcome(s) of interest across the control and experimental cases using paired t-tests.

Table 6. Precision Matching: Comparison of Covariates for Prison and Community Control (C.C.) Groups Using No Matching and Precision Matching (Bales and Piquero, 2012)

	No Matching (n=144,416)		Matching (n=18,712)	
	Prison (79,022)	C.C. (65,394)	Prison (9,356)	C.C. (9,356)
Male	.915	.813*	.890	.890
White	.377	.508*	.448	.448
Black	.564	.406*	.492	.492
Hispanic	.059	.086*	.060	.060
Age	32.460	30.441*	31.005	31.005
Murder/Mansl.	.015	.006*	.013	.013
Sex	.033	.040*	.044	.044
Robbery	.079	.037*	.047	.047
Other Violent	.149	.177*	.173	.173
Burglary	.180	.130*	.142	.142
Property	.141	.175*	.158	.158
Drug	.291	.309*	.317	.317
Weapon	.036	.034	.033	.033
Other	.076	.092*	.073	.073
Prior Felonies	1.654	1.070*	.512	.512
Prior Prisons	1.253	.410*	.589	.589
Prior Super. Viol.	1.388	.857*	.988	.988
Offense Level	5.249	4.564*	4.950	4.950
Recomm. Prison	.756	.378*	.551	.551

* p<.001.

Table 7. Precision Matching: Recidivism Outcomes of Prison and Community Control (C.C.) Groups Using No Matching and Five Matching Models (Bales and Piquero, 2012)

	Recidivism-1 Year			Recidivism–2 Years			Recidivism-3 Years		
	Prison	C.C.	Diff.	Prison	C.C.	Diff.	Prison	C.C.	Diff.
No Matching (n=144,416)	27.0%	15.3%	11.7%	40.0%	24.8%	15.2%	47.8%	32.4%	15.4%
Matching #1 (n=12,088)	9.3%	7.4%	1.9%	14.7%	12.0%	2.7%	18.2%	15.4%	2.8%
Matching #2 (n=15,714)	17.4%	13.4%	4.3%	26.3%	20.6%	5.7%	32.0%	26.4%	5.6%
Matching #3 (n=19,840)	24.4%	17.2%	7.2%	36.2%	27.4%	8.8%	43.6%	35.4%	8.2%
Matching #4 (n=20,314)	27.2%	18.1%	9.1%	40.2%	28.8%	11.4%	48.1%	37.6%	10.5%
Matching #5 (n=18,712)	28.1%	18.5%	9.6%	41.3%	29.4%	11.9%	49.1%	38.4%	10.7%

Notes: All Mean differences are significant at p<.001.
Matching #1: Sex, Race, Age, Current Offense, and Prior Convictions.
Matching #2: Sex, Race, Age, Current Offense, Prior Convictions. And Prior Prison Commitments.
Matching #3: Sex, Race, Age, Current Offense, Prior Convictions, Prior Prison Commitments, and Prior Supervision Violations.
Matching #4: Sex, Race, Age, Current Offense, Prior Convictions, Prior Prison Commitments, Prior Supervision Violations, Current Offense, and Sentencing Guidelines Level.
Matching #5: Sex, Race, Age, Current Offense, Prior Convictions, Prior Prison Commitments, Prior Supervision Violations, Current Offense, Sentencing Guidelines Level, and Scored to Prison.

COARSENED EXACT MATCHING (CEM)

- Developed recently by Iacus et al. (2011) as an alternative to exact matching.
- The CEM matching algorithm allows the user to conduct exact matching in addition to many different approaches to coarsening the data.
- Coarsening is the process by which a continuous, ordinal, or even multi-category nominal variable is segmented into a reduced set of strata.
- Coarsening a variable such as age into a set of discrete categories reduces the dimensionality of the matching process and will yield more matches.
- The causal analysis is based only on the matched observations. Both the matching process and causal analysis are transparent.

COARSENED EXACT MATCHING(CEM)-CONTINUED

- To date, coarsened exact matching has not been applied to substantive criminal justice issues, yet has several attractive features.
- CEM is relatively straightforward and easy to apply in statistical software packages including Stata, SPSS, and R.
- The matching procedure originates from a class of matching methods that circumvents the sometimes arduous balance-checking phase when using PSM.
- The CEM procedure is efficient in samples or populations with large N sizes.
- This approach allows for the estimation of a multivariate imbalance statistic (L1) that calculates the overall between-group imbalance among all covariates as well as the level of imbalance for each specific covariate. L1 = 0 is perfect balance, L1 = 1 is no balance. L1 = .475 indicates 52.5% (1-.475) balance, or overlap on all of covariates across the two group.
- Easy to examine differences between matched and unmatched cases.

Coarsened Exact Matching (CEM)- Stata Code to conduct exact matching

- Sample Stata code in which we are conducting Exact Matching using CEM to create equivalent cases who were sentenced to prison versus non-prison to determine the effect of imprisonment on the likelihood of recidivism within 2 years.

- cem male (#0) age_at_sentencing (#0) curr_off_5_groups_violent (#0) curr_off_5_groups_sex (#0) curr_off_5_groups_property (#0) curr_off_5_groups_drug (#0) curr_off_5_groups_other (#0) black (#0) hispanic (#0) priors_prison_commitments (#0), tr(sanction_prison)

- ttest re_conviction_within_2_years, by(sanction_prison)

Coarsened Exact Matching (CEM)- Stata Code to conduct cem

- Sample Stata code in which we are conducting CEM to create equivalent cases who were sentenced to prison versus non-prison to determine the effect of imprisonment on the likelihood of recidivism within 2 years.
- The code is letting CEM determine the level of coarsening of the variables to reach optimal balance between the control and experimental groups.
- cem male age_at_sentencing curr_off_5_groups_violent curr_off_5_groups_sex curr_off_5_groups_property curr_off_5_groups_drug curr_off_5_groups_other black hispanic priors_prison_commitments, tr(sanction_prison)
- logit re_conviction_within_2_years sanction_prison [iweight=cem_weights]

Coarsened Exact Matching (CEM)- Stata Code to conduct cem With groupings of continuous variables

- If you have one or more continuous variables, such as age, number of prior convictions, etc. in which you want to specify the groupings of the values, you can do this in CEM.

- The code below is specifying the groupings for age at sentencing and prior prison commitments.

- cem male age_at_sentencing (10 17 21 29 40 99) curr_off_5_groups_violent curr_off_5_groups_sex curr_off_5_groups_property curr_off_5_groups_drug curr_off_5_groups_other black hispanic priors_prison_commitments (1 3 5 10), tr(sanction_prison)

- logit re_conviction_within_2_years sanction_prison [iweight=cem_weights]

Advantages and disadvantages of the three matching methods

- Propensity Score Matching (PSM) Advantage – In studies with smaller N sizes, PSM is more appropriate than Exact or CEM matching.
- Propensity Score Matching (PSM) Disadvantage – The process of balancing can be time consuming and may require truncating variables and losing specificity in your measures.
- Exact Matching Advantage – Balancing is not required and the coding required using CEM is very simple. Also, this method is easy to understand for non-researchers.
- Exact Matching Disadvantages - Requires large N sizes and the attrition of a significant number of cases may result, which calls into question the generalizability of the results.
- Coarsened Exact Matching (CEM) Advantages - Relatively easy to use and flexible in terms of specifying the level of precision of matching you want to achieve.
- Coarsened Exact Matching (CEM) Disadvantages - Requires large N sizes.

References

Bales, W. D., & Piquero, A. R. (2012). Assessing the impact of imprisonment on recidivism. *Journal of Experimental Criminology*, *8*, 71–101.

Blackwell, M., Iacus, S. M., King, G., & Porro, G. (2009). CEM: Coarsened exact matching in Stata. *The Stata Journal*, *9*, 524–546.

Dirkzwage, Anjua, King, R. , Massoglia, M., & Macmillion, R. (2007). The context of marriage and crime: Gender, the propensity to marry, and offending in early adulthood, C*riminology*, 45, 33-55.

King, R. , Massoglia, M., & Macmillion, R. (2007). The context of marriage and crime: Gender, the propensity to marry, and offending in early adulthood, C*riminology*, 45, 33-55.

Iacus, S. M., & King, G. (2012). *How coarsening simplifies matching-based causal inference theory* (vol. 15). New Brunswick, NJ: Transaction Pub.

Iacus, S. M., King, G., & Porro, G. (2011a). *Causal inference without balance checking: coarsened exact matching, Political Analysis Advance Access*. Oxford, England: Oxford University Press.

Iacus, S. M., King, G., & Porro, G. (2011b). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*. Retrieved from http://gking.harvard.edu/files/abd/CEM-math-abs.shtml.

Mears, Daniel P. and William D. Bales. (2010). Supermax Housing: Placement, Duration, and Time to Reentry. *Journal of Criminal Justice*, 38, 545-554.

Nagin, Daniel S., Francis T. Cullen, and Cheryl Lero Jonson. (2009). Imprisonment and reoffending. In Michael Tonry (Ed.), *Crime and Justice: A Review of Research, Volume 38* (pp. 115-200). Chicago: University of Chicago Press.

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