Comparison Groups for Treatment Court Outcome Evaluations

Dana Hurley Swayze, MSW
Matthew A. Johnson, Ph. D.

Presented at the NADCP National Conference
Washington DC, 2017
Presentation Overview

• Roles of audience members and presentation goal
• Background about Hennepin County and its Treatment Courts
  • Recent evaluations of Drug Court and Veterans Court
• Importance of comparison groups in outcome evaluations
• Pitfalls to constructing comparison groups
• Propensity score matching technique to the rescue
• Experience of constructing comparison groups from recent evaluations:
  • Drug Court
  • Veterans Court
• Lessons Learned
• Questions
Presentation Goals

- Understand the lessons learned from Hennepin County’s experience assembling comparison groups for two different treatment court outcome evaluations
- Demonstrate how to use matching techniques to minimize bias and/or dissimilarity in comparison groups
- Demonstrate the importance of high levels of cooperation among treatment court team members for creating comparison groups (and high quality outcome evaluations in general)
MN and Hennepin County

- Minnesota’s population is approx. 5.5 million
- Hennepin County is the largest in terms of population in the state (1.2 million; 22%)
  - Minneapolis 1/3 of County population
- Minnesota has 87 counties divided into 10 Judicial Districts
- The 4th Judicial District is geographically synonymous with Hennepin County
- Hennepin Co. includes the city of Minneapolis and 46 surrounding communities
- 4th Judicial District handles the most case filings in MN (~500,000 in 2016)
Hennepin Treatment Courts

Fourth Judicial District has 4 Treatment Courts:
• Model Drug Court (1997); ~130 participants
• Mental Health Court (2004); ~175 participants
• DWI Court (2007); ~125 participants
• Veterans Court (2010); ~60 participants

• All treatment courts are voluntary on the part of the participant in lieu of the stayed sentence
• Some courts are post-adjudication while others allow admission before a criminal case is fully resolved
Comparison Groups: Why?

• How do we know if programs, such as treatment courts, are meeting their goals? Do treatment courts function better than “business as usual”?

• We may assume treatment court programming is leading to better outcomes, but unless outcomes are compared to similarly-situated criminal offenders that did not participate in the program, it is impossible to draw firm conclusions about the effectiveness of treatment courts
  • Is a treatment court recidivism rate of 25% good? Even if the general recidivism rate is 30%, it is not sufficient to say that treatment court participants have better outcomes because we are dealing with a specific subpopulation

• Goal is to compare apples to apples (or as close to apples as possible...there will always be variance we cannot model)
NADCP Practice Standard

Monitoring and Evaluation, Comparison Groups

• Outcomes for Drug Court participants are compared to those of an unbiased and equivalent comparison group.

• Individuals in the comparison group satisfy legal and clinical eligibility criteria for participation in the Drug Court, but did not enter the Drug Court for reasons having no relationship to their outcomes.

• Comparison groups do not include individuals who refused to enter the Drug Court, withdrew or were terminated from the Drug Court, or were denied entry to the Drug Court because of their legal charges, criminal history, or clinical assessment results.
NADCP Practice Standard

- **Random Assignment**: Truly random assignment to the treatment or control group
- **Quasi-experimental Comparison Group**: Using a wait-list cohort, or a neighboring county without a treatment court, or a time period prior to a treatment court as a comparison group
- **Matched Comparison Group**: Constructing a comparison group out of a large pool of offenders, such as a probation database
- **Propensity Score Analysis**: Advanced statistical matching procedure to calculate the statistical probability that an individual with a given set of characteristics would be in the treatment group as opposed to in the comparison group. (The relative similarity of that one individual to one group as opposed to the other)
- **Invalid Comparison Groups**: Those who weren’t accepted, refused, terminated, or are in some other way biased
Propensity Score Matching

- Random assignment is very rare in treatment courts, so treatment groups need to be matched to a comparison group.
- Matching groups on a single element (e.g. gender) is rather simple, and can be done without specialized programs.
- There is a problem when trying to match on multiple covariates simultaneously (gender and race and age and criminal history, etc.).
- Propensity score matching can help overcome the difficulties of matching treatment and control groups.
  - “Rather than requiring close or exact matches on all covariates, [propensity score matching] enables the construction of matched sets with similar distributions of covariates”

- Stuart and Rubin (2007)
Propensity Score Matching

• In short, propensity score matching allows the creation of a comparison group that is similar to the control group while minimizing potential selection bias
  • Pros:
    • Flexibility
      • Multiple matching techniques, flexible distance measure (a.k.a. caliper)
    • More reliable than the human eye
  • Cons:
    • Propensity score matching packages in statistical programs are not all created equal (be very wary of SPSS’s built-in PSM!)
    • Can’t model every aspect of both groups, so there can be hidden bias
    • Dependent on quality data (garbage in, garbage out)
Propensity Score Matching

• BUT...propensity score matching is no replacement for the researcher.

• Requires analysis of outcome data after matching
  • Are the groups systematically matched across all variables?
  • Are there any statistically significant differences between the groups on any variables?
    • If so, does extant the literature suggest if these variables impact criminal justice outcomes?

• A case study of the built-in propensity score matching feature in SPSS
  • Few options and can return matches that differ in significant ways

• In short, a great tool, but not a silver bullet on its own
Propensity Score Matching

• How Hennepin County used propensity score matching in its Drug and Veterans Court evaluations:
  • Matched on race/ethnicity, age, gender, criminal history, and instant offense

• Drug Court:
  • 317 in evaluation cohort; 620 potential matches
    • About a 2-1 ratio
  • Used 1:1 matching
    • Couldn’t find a match for 18 Drug Court participants (6%)

• Veterans Court:
  • 155 in evaluation cohort; 220 potential matches
    • Less than 1.5:1 ratio
  • Used 1:many matching
    • Matched all VC participants to 82 unique individuals in comparison group
Propensity Score Matching

• Ultimately, the matched comparison groups for both treatment courts were statistically identical to the evaluation cohorts based on the included covariates
  • But this only came after a couple of false starts
• Each treatment court had to consider the various tradeoffs:
  • 1:1 matching could not find a match for every individual in the Drug Court cohort
  • 1:many matching led to far fewer unique individuals in the Veterans Court comparison group
• What’s right for you? Depends on the size of your cohort and the number (and quality) of potential matches
Model Drug Court

- Established in 1997; revamped in 2007 to come into alignment with national best practice standards
- MDC uses interdisciplinary staffing team and steering committee, as well as one judge model
- Presently a four stage program, though it was three stages during the evaluation period
- Capacity is 135 participants under supervision of four probation agents
- Currently an 18 month program
Model Drug Court

Admission Criteria

• Hennepin County resident

• Felony Level Drug Offense –or—
• Felony Level Property Offense deemed related to a drug habit

• Must be high risk and high need to reoffend based on the Risk and Need Triage assessment (RANT) administered by probation staff

• Must have a diagnosed substance use disorder based on Minnesota’s Rule 25 CD Assessment or an equivalent private assessment. Current DSM criteria.

Exclusionary Criteria

• History of violent offenses; use of gun in current offense; MN DOC parolee; part of a drug manufacture/distribution network; mental incompetency
Model Drug Court

- Evaluation cohort consists of those who completed MDC (successfully or unsuccessfully) between 2011 and 2013
- Allows for an additional 2 years after MDC to explore recidivism (2013 to 2015)
- Allows us to extend the recidivism window out further for those who spent any time incarcerated to have two full years of “street time” (into 2016)
- No prior experience with MDC or any other Henn. Co Treatment Court
- Starting cohort was 317 individuals
Prepping a Match

• 4th Judicial Research collaborated with the Hennepin Co. Dept. of Community Corrections and Rehabilitation (DOCCCR) to procure a data set on adults under supervision in the county (2008-2013)
• Goal to identify who potentially could have been in MDC
• Initial file was 32,556 records: duplicate cases and clients!
• First, we needed an equivalent for high risk and high need to reoffend
• Minnesota probation services use the Level of Service-Case Management Inventory (LS-CMI) to establish if clients are high risk to reoffend. We kept probationers in our sample who scored in the high risk or very high risk category on the LS-CMI.
Prepping a Match

- We also needed a proxy for a **substance use disorder** in the absence of CD diagnostic information in the probation dataset (data privacy issues)
- There are numerous risk “domains” in the LS-CMI, so we selected only probationers who scored **high risk** in the substance use disorder domain of the tool (as well as high risk on the tool overall)
- While the LS-CMI is a much less comprehensive evaluation of substance use disorders than a Rule 25, an LS-CMI is a much more comprehensive assessment of risk and need than the RANT tool
- The assessment tools are not “apples to apples” across the two cohorts, but you do the best you can to mitigate differences!
Prepping a Match

Other Considerations

• Using their unique court party ID we could exclude them if they had ever been referred or accepted to any other Problem Solving Court in MN—no prior treatment effect!
• Using case and disposition information we could ensure they were on probation for a felony level drug or felony property offense
• Ensured probation was related to a Hennepin County case
• Ensured probation was post-disposition
• Selected comparable level of administrative or community supervision
• The latter are important because you want the same suite of probation tools/sanctions to apply to your respective cohorts based on their disposition status and level of monitoring
The Chosen Few

- Ultimately, 620 probationers met criteria to potentially be matches for the 317 MDC participants.
- Serious attrition occurred while winnowing the pool down to offenders with like-characteristics, and this precedes any matching on demographics and criminal history!
- We were now ready to use Propensity Score Matching to find out which of the 620 made the highest quality matches to the MDC participants.
- Matches were ultimately based on: age, gender, offense (drug or property), race (white, non-white), and a prior criminal history score variable.
- The PSM package set at .05 caliper resulted in 299 matches, meaning 18 MDC participants were lost from the evaluation (6%). The tighter the caliper, the better the match quality but the fewer matches are made.
**Pre-Post PSM Comparison**

<table>
<thead>
<tr>
<th>Populations Before Propensity Score Matching</th>
<th>MDC N=317</th>
<th>DOCCR N=620</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female*</td>
<td>34%</td>
<td>13%</td>
</tr>
<tr>
<td>Person of Color*</td>
<td>57%</td>
<td>69%</td>
</tr>
<tr>
<td>Under Age 25*</td>
<td>19%</td>
<td>26%</td>
</tr>
<tr>
<td>Age 41 or Older*</td>
<td>35%</td>
<td>24%</td>
</tr>
<tr>
<td>Instant Offense: Drug</td>
<td>67%</td>
<td>68%</td>
</tr>
<tr>
<td>No Prior Convictions</td>
<td>10%</td>
<td>14%</td>
</tr>
</tbody>
</table>

**Populations After 1:1 Propensity Score Matching**

<table>
<thead>
<tr>
<th>MDC N=299</th>
<th>DOCCR N=299</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>30%</td>
</tr>
<tr>
<td>Person of Color</td>
<td>58%</td>
</tr>
<tr>
<td>Under Age 25</td>
<td>19%</td>
</tr>
<tr>
<td>Age 41 or Older</td>
<td>36%</td>
</tr>
<tr>
<td>Instant Offense: Drug</td>
<td>67%</td>
</tr>
<tr>
<td>No Prior Convictions</td>
<td>10%</td>
</tr>
</tbody>
</table>

**MDC cohort and DOCCR population are statistically different at \( p \leq 0.001 \)**

MDC and DOCCR are not statistically different on any variables
Veterans Court

- Started in June 2010
- Veterans Court uses interdisciplinary staffing team and steering committee, as well as one judge model
- Largest Veterans Court in Minnesota, current census about 60
- Program admission criteria:
  - Veterans of US Armed Forces (very broadly defined) charged with a criminal offense in Hennepin County
    - Program rules include some specific instant offense and criminal history exclusion criteria
  - Ideal comparison group would be a cohort of individuals in the criminal justice system who had served in the military, but who were never referred to the Hennepin County Veterans Treatment Court
Veterans Court

• Program report after two years, but no comparison group of justice-involved veterans included
  • Just examined “outcomes” of participants before and after admission to court
• ROADBLOCK: the Minnesota Court Information System did not capture veteran status of justice-involved veterans, nor did any justice partners on the Veterans Court team
  • No readily available way to discern veterans in criminal justice system who were not referred to the program
Veterans Court

• SOLUTION: Work with VA to find offenders in the criminal justice system who were similar to the 155 Veterans Court participants being evaluated, but had not been referred to the program (or any of Minnesota’s treatment courts)

• Required teamwork:
  • Research department would look for potential matches
  • VA would determine veteran status of the potential matches

• Research department created certain offender profiles based upon the criminal histories and instant offenses of Veterans Court participants; 21 different profiles
  • For example, Gross Misdemeanor DWI offenders with no prior convictions or 5th Degree Assault offenders with more than three prior convictions
Veterans Court

• Pulled data from Court Information System to identify as many Hennepin County offenders as possible who matched the profiles of the Veterans Court cohort
  • Resulted in anywhere from 50 to hundreds of potential matches, depending on the different instant offense and criminal history combinations
  • Told VA how many veterans to find in each group to end up with 155 justice-involved veterans who had not been referred to the Veterans Court program
Veterans Court

• Data sharing court order allowed the court to transfer offender data to the VA
  • The VA sent back data that did not include any sensitive information, in accordance with their rules
• Without the court order, creating a comparison group of justice involved veterans would have been impossible
Veterans Court

- VA looked up each potential match one-by-one to determine which individuals had verified military service histories
  - Time consuming and tedious task!
  - Demonstrates importance of strong and willing treatment court team members
- VA team returned 155 individuals as requested, BUT...
- There were systematic differences between the Veterans Court evaluation cohort and the comparison group
  - Initially didn’t include race, age, and gender in the profiles of the Veterans Court evaluation cohort
  - Comparison group was significantly younger than the Veterans Court evaluation cohort, and the Criminal Justice literature has consistently demonstrated age is related to criminogenic activity
Veterans Court

• The Research Department sent another list of potential comparison group members to the VA who were all over 40
  • VA graciously agreed to find additional veterans from the new list
• All told, the VA identified around 220 individuals whose military history could be verified, BUT...
• How to match the comparison group of justice-involved veterans to the Veterans Court evaluation cohort?
Veterans Court

• Ultimately, we settled on a one-to-many propensity score matching strategy to create the comparison group
  • Doing a one-to-one match with a reasonable caliper resulted in too many individuals from the Veterans Court evaluation cohort without a match
  • Faced with a tradeoff between fewer individuals to evaluate and a 1:1 comparison group, we opted to maximize the number of Veterans Court participants being evaluated
• Matching on a one-to-many basis resulted in a comparison group of justice involved veterans that was statistically identical to our Veterans Court evaluation cohort based on age, race/ethnicity, gender, criminal history, and type of instant offense
### Veterans Court

<table>
<thead>
<tr>
<th></th>
<th>Veterans Court Evaluation Cohort</th>
<th>Full Potential Comparison Group</th>
<th>Comparison Group after PSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>3.2%</td>
<td>4.6%</td>
<td>2.6%</td>
</tr>
<tr>
<td>Non-While</td>
<td>34.2%</td>
<td>43.1%</td>
<td>41.2%</td>
</tr>
<tr>
<td>Age</td>
<td>43.4</td>
<td>44.0</td>
<td>44.7</td>
</tr>
<tr>
<td>Criminal History Points (mean)</td>
<td>4.5</td>
<td>6.9</td>
<td>5.3</td>
</tr>
<tr>
<td>Current Charge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Other Non-Traffic</td>
<td>15.5%</td>
<td>18.98%</td>
<td>13.6%</td>
</tr>
<tr>
<td>• Misdemeanor DWI</td>
<td>5.2%</td>
<td>7.87%</td>
<td>5.2%</td>
</tr>
<tr>
<td>• 5th Degree Assault</td>
<td>14.8%</td>
<td>14.35%</td>
<td>14.8%</td>
</tr>
<tr>
<td>• Other Gross Misdemeanor</td>
<td>9.7%</td>
<td>9.26%</td>
<td>9.0%</td>
</tr>
<tr>
<td>• Gross Misdemeanor DWI</td>
<td>35.5%</td>
<td>28.24%</td>
<td>34.8%</td>
</tr>
<tr>
<td>• Other Felony</td>
<td>19.4%</td>
<td>21.3%</td>
<td>22.6%</td>
</tr>
</tbody>
</table>
Veterans Court

• Known limitations of comparison group:
  • No military service information
    • No rank information, No branch of service data, No discharge data, No deployment data
  • No VA service/benefit connectedness information
  • No substance use disorder or mental health history
  • No treatment information
  • No employment, housing, or education data
  • Participant race data was dichotomized to white/nonwhite, which might obscure different program effects between nonwhite groups (maybe Asian participants fare differently than Latinos)
• But we knew we were comparing program participants to individuals who met Veterans Court criteria
Veterans Court

- If this seems like a lot of work...it was.
- Question: do you need to a comparison group with verified military history to have a valid comparison group for a Veterans Court evaluation?
  - In a follow-up analysis conducted after the completion of the evaluation, we tested this question.
  - The results have implications for other courts looking to properly evaluate the outcomes of their Veterans Treatment Court participants
    - That is, do you need to go through all the work of identifying veterans who were not referred to the program?
Key Lessons Learned

• Need to be thinking about evaluations YEARS IN ADVANCE
  • Need to consider what data you’ll need
  • Need to get participant consent forms in place
  • Need to get and data sharing agreements in place
    • Need to know what potential barriers to sharing or receiving information exit

• Evaluations require a coordinated effort and a unified team willing to do whatever it takes
  • How well the team works together will greatly influence the availability and quality of participant data
  • Establish roles and expectations before evaluation process starts
Key Lessons Learned

• Need to know what is distinct about participants entering your treatment court
  • Constructing proper comparison group contingent on this
  • For example, requiring participants to be high-risk, high-need means you need to compare them against similarly situated offenders

• Know your data! If something seems off, it probably is
  • Be willing to say things are wrong and go back to the drawing board if necessary
    • Better to spend more time preparing quality data than having to defend poor data
Key Lessons Learned

• Data availability vs. reality
  • Tradeoffs between one-to-one matches vs. one-to many matches
  • Are there defendable explanations for limitations?
  • Is an imperfect comparison group better than a non-existent comparison group?
    • Hint: it likely is!

• Technology is important for constructing comparison groups, but it must be paired with an evaluator’s trained eye
  • Propensity score matching is merely a tool. Whether it’s doing its job properly needs to be assessed by someone who knows the data well (a.k.a. you!)
Thank you!

- Questions???