Published in final edited form as:

Accid Anal Prev. 2013 April ; 53: 112–120. doi:10.1016/j.aap.2012.12.033.

Effects of Admission and Treatment Strategies of DWI Courts on Offender Outcomes

Frank A. Sloan^{a,b}, Lindsey M. Chepke^a, Dontrell V. Davis^a, Kofi Acquah^{a,c}, and Phyllis Zold-Kilbourne^d

^aDuke University, Department of Economics, Box 90097, Durham, NC 27708

^dBuy The Numbers, Inc. 6151 Cold Springs Trail, Grand Blanc, MI 48439

Abstract

Purpose—The purpose of this study is to classify DWI courts on the basis of the mix of difficult cases participating in the court (casemix severity) and the amount of involvement between the court and participant (service intensity). Using our classification typology, we assess how casemix severity and service intensity are associated with program outcomes. We expected that holding other factors constant, greater service intensity would improve program outcomes while a relatively severe casemix would result in worse program outcomes.

Methods—The study used data from 8 DWI courts, 7 from Michigan and 1 from North Carolina. Using a 2-way classification system based on court casemix severity and program intensity, we selected participants in 1 of the courts, and alternatively 2 courts as reference groups. Reference group courts had relatively severe casemixes and high service intensity. We used propensity score matching to match participants in the other courts to participants in the reference group court programs. Program outcome measures were the probabilities of participants': failing to complete the court's program; increasing educational attainment; participants improving employment from time of program enrollment; and re-arrest.

Results—For most outcomes, our main finding was that higher service intensity is associated with better outcomes for court participants, as anticipated, but a court's casemix severity was unrelated to study outcomes.

Conclusions—Our results imply that devoting more resources to increasing duration of treatment is productive in terms of better outcomes, irrespective of the mix of participants in the court's program

Keywords

Alcohol-impaired driving; Drunk Driving; Court Intervention; DWI offenders; Motor vehicle crashes; Treatment

© 2012 Elsevier Ltd. All rights reserved.

^bCorresponding Author: Duke University, Box 90097, Durham, NC 27708, fsloan@duke.edu, Phone: 919-613-9358, Fax: 919-681-7984.

^cPresent Address Brown University, 69 Brown Street, Economics Box B Providence, Rhode Island, 02912

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1.0. INTRODUCTION

Externalities from drinking and driving behaviors are well documented. While there has been a dramatic decrease in the number of alcohol related fatal crashes over the last decade, the total number of fatal crashes overall has decreased as well (National Highway Traffic Safety Administration, 2009). Nationally, the percentage of alcohol-impaired driving fatalities has remained at 32%; between 2000–2009, the percent of alcohol-impaired passenger vehicle drivers involved in fatal crashes remained practically unchanged (National Highway Traffic Safety Administration, 2009, 2011). Given the externalities from reckless driving, governments have enacted and enforced laws to promote safe driving, invested in roadways, promulgated regulations to promote vehicle safety, controlled entry of alcohol sellers, and other laws directly aimed at discouraging driving while intoxicated (DWI).¹ A more recent, though certainly not new, policy approach is implementation of DWI treatment courts.

Modeled on the concept of drug courts, DWI courts have been established throughout the U.S. to integrate penalties for DWI violations with treatment for underlying alcohol addiction. A rationale for DWI courts is that the conventional traffic court model of prosecuting DWI offenders fails to address the addiction component of DWI cases. Because treatment courts utilize a variety of services, many of these courts are tailored to deal with more than one dimension of DWI behavior.

There have been several evaluations of DWI and hybrid courts' effectiveness (Eibner et al., 2006, MacDonald et al., 2007, Moore et al., 2008, Bouffard et al., 2010, Bouffard and Bouffard, 2011). In general, the outcome measure is re-arrest for DWI or for any offense. Based in part on unpublished studies, the general conclusion from existing evidence is that DWI/ hybrid courts deter repeat offenses (Huddleston and Marlowe, 2011). Nevertheless, not all findings are entirely positive. Bouffard et al. (2010) found that DWI courts are ineffective in preventing recidivism for individuals previously charged with a DWI but are effective for individuals with non-DWI offenses. In a study on the cost effectiveness of a California DWI court, Eibner et al. (2006) concluded that DWI court was less costly than traditional courts for third-time offenders, but more costly for second time offenders, which implies that the DWI court was comparatively more effective in dealing with hard-core offenders. Irrespective of the prior offense record, the authors concluded that the DWI court improved offender outcomes across several measures (e.g., alcohol problems index, drinks per day, stressful life events index, recidivism).

Many of the studies suffer from important methodological limitations. Some studies lack a control group or sufficient evidence that the control group adequately matches the DWI court group. Sample sizes are often so small that the studies are underpowered. The vast majority of studies are based on localized samples. Also, positive results may be much more likely to be published or more generally, to be publicized (publication bias).

Like the previous studies, we conducted this study to assess the effectiveness of DWI courts with varying casemixes and service intensities on program outcomes. We defined casemix severity as the mix of difficult cases participating in the court and service intensity as the amount of involvement between the court and participant. We hypothesized that for a given level of service intensity, courts with a mix of participants for which favorable program outcomes are inherently more difficult to achieve (higher casemix severity) are would in fact have poorer outcomes since coping with a more severe casemix would place higher demands

¹For ease of reading, we refer to drinking and driving generally as DWI, regardless of what an individual state calls this offense. This term varies by state and jurisdiction and is also called e.g., DUI, OWI, OUI, OMVI, DUIL, DWAI, or DWUI.

on a fixed quantity of program resources. Other factors being equal, we hypothesized that a greater amount of services per court participant (higher service intensity) would lead to better outcomes. In other words, we tested for whether or not a greater expenditure of resources in fact produced better program outcomes. Our study improved on previous studies specifically, by having an adequate control group and sample size, and in using propensity score matching (PSM) to control for the heterogeneity of individual DWI treatment courts.

In the next sections, we provide an overview of our statistical methods, describe the courts in our sample, describe our statistical methods in detail, and then present a review of our findings. Finally, we discuss our findings in the context of what is known about DWI and hybrid courts as well as our conclusions.

2.0. METHODS

2.1. Overview

We classified courts based on their casemix severity and service intensity. Casemix severity was determined by several characteristics of court participants, mental health history, use of scheduled 1 or 2 controlled substances, or prior substance abuse treatment. Service intensity was based on the requirements of each court, e.g., days of treatment required or use of sanctions and incentives. Before matching, we classified courts on the basis of the mix of difficult cases participating in the court (casemix severity) and the amount of involvement between the court and participant (service intensity). Then after matching, we compared matched groups on the basis of their program outcomes. We use propensity score matching (PSM) to account for the differences in our sample. PSM involves the following steps. First, logit analysis is conducted to assess correlates of a person being in the treatment reference versus the control group. In our study, since participants were “treated” in all sample courts, we refer to the what is typically called the “treatment” group as the “reference” group and the control group is referred to as the comparison court sample. Second, using the parameter estimates from the logit regression, a predicted probability of being a participant in the reference court sample is calculated, both for participants in the reference court program and (counterfactually) in the comparison court program. These predicted probabilities are used for matching. This method of PSM pairs reference court participants with comparison court participants whose propensity scores (probabilities of being a participant in the reference court program) differ up to a pre-specified amount (this is known as the caliper width) (Austin, 2011). We used one to one nearest neighbor matching without replacement using a 0.05 caliper. Finally, once the groups are matched, an average treatment effect (ATT) is calculated for each matched pair of participants. The ATT compares outcomes of participants in the reference court(s) with those in the comparison court samples.

2.2. Data

Data came from seven DWI courts in Michigan and one DWI court in North Carolina. We included two states in our analysis to provide results that are not unique to the idiosyncrasies of an individual state’s DWI court program. Michigan was chosen for two main reasons; it has a well-established DWI court program and has programs in various geographic areas of the state. As of 2010, Michigan had 24 designated DWI courts (Michigan Drug Treatment Courts, 2010). Individual Michigan DWI courts were selected based on their prior participation in DWI research and willingness to share their data. North Carolina was chosen because of the quality and availability of their court data. At the time this study was conducted, North Carolina had one operating DWI court.² The data used in our analysis spanned 2004 to 2010 for Michigan and 2000 to 2010 for North Carolina. In total, our eight-

court sample contained 3,844 observations. Analysis of individual court pairs differed in numbers of observations due to missing values.

Both Michigan and North Carolina maintain DWI court data at the court level while general criminal court data is maintained at the state level. North Carolina criminal court data were provided for the entire state and includes DWI arrest data for all 100 counties. Due to restrictions in Michigan data policies, criminal court data were only obtained for the seven counties with which we had existing data use agreements with DWI courts. As a result, our recidivism analysis is limited to arrests occurring in the seven Michigan counties used in our study. This limitation is mitigated by the fact that most persons are arrested for DWI in their county of residence; a minority of rearrests would occur in counties outside of where the initial arrest occurred. To support this, in a separate analysis of North Carolina criminal court data for 2001–2011 from all 100 counties in the state we found that 70 percent of arrests for DWI occurred in the arrestee's county of residence. This provides a rough estimate of the share of total DWI arrests we could measure in Michigan.

The court data contained information on substance abuse and mental health history at entry, information on demographic characteristics of program participants, and several measures of outcomes, including whether or not the participant completed the program, and whether or not there were changes in educational attainment and in employment status between the date of entry and the time the individual completed the program. We also obtained information on DWI arrests and convictions in the counties in which the DWI courts were located in Michigan and for the entire state of North Carolina. In accordance with our data agreement with the courts, each court was assigned a label corresponding to its position in terms of service intensity and the severity of cases it accepted (see Figure 1). The mean values of the casemix and intensity indexes were 0.40 and 0.00, respectively, the latter having been normalized to zero. The casemix index varied from 0.20 to 0.45; the intensity index varied from –1.5 to 5.5.

The data formats and content were the same for the Michigan courts in the sample. While the format and content differed between North Carolina and Michigan, the variables used in our empirical analysis were comparable between the two states. However, there were a few differences in data between the states. For example, the North Carolina court did not obtain information on changes in educational attainment or recidivism. Courts in Michigan obtained data on recidivism, but several courts reported rates of re-arrest for any offense including DWI well below 1%. For at least 1 DWI court, the low re-arrest rate reflected a short follow-up period. In some cases, such rates could reflect poor ascertainment of rearrests, and we did not analyze data with re-arrest rates from Michigan courts that reported rates below 1% for this reason.

Enrollment criteria for DWI offenders varied by DWI court, however, there were similar requirements across courts. Participants must: be deemed to have a substance dependency, a DWI charge, residence in the county where the court is located, and no prior violent crimes or felonies.³ All cases in the court data were included in our analysis, with the exceptions noted below in the results section. Additionally, while some courts had more cases, this is reflective of the size of the DWI treatment court, not our analysis.

Our sample of participants in Michigan DWI courts was representative of DWI courts in Michigan as of 2009–2010 in terms of important published attributes. In Michigan, 60% of

²As of 2012, The Mecklenburg STEP court is the only active DWI court in North Carolina. However, this court is divided into two separate courts with one court being primarily Spanish speaking. For purposes of our analysis, we have treated the courts as a single court.

³The eligibility criteria were gathered from participant handbooks from the individual DWI courts.

participants completed the program statewide; compared with a 61% completion rate in our sample (comparison of data in Michigan Drug Treatment Courts (2010) with our data). In both the universe of Michigan DWI courts and in our sample, 9% of participants at entry were black; our sample contained 9% Hispanics versus 6% for DWI courts in Michigan overall. Persons of other race/ethnicity were 3% in both samples. Our sample consisted of 24% females versus 27% for Michigan DWI courts overall. In both samples, median educational attainment at entry was 12 years. 25% of our sample was not employed (either not in the labor force or unemployed at entry to the program); compared with a rate of 32% not employed for Michigan DWI courts overall.

2.3. Classifying Courts

The courts differed in both the types of cases they accepted and in the intensity of treatment. We developed a court-specific casemix index as follows. First, we specified and, using data from all court participants in the entire sample, we estimated an equation using logit analysis with the individual program participant as the observational unit. The dependent variable was a binary variable set equal to 1 if the person did not complete the program and was set to 0 otherwise. Explanatory variables measured substance abuse and mental health history at entry into the program and demographic characteristics. Mental health history was defined as a binary variable that measured whether an individual had a medical diagnosis from the Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV) before entry into the court. Substance abuse was defined by the court as to whether or not the individual: had any prior substance abuse treatment; had any prior experience with intravenous drugs; used drugs or alcohol before age 16; the person's main addiction was to alcohol, either reported as such or inferred from a DSM-IV code; the person used Schedule 1 (e.g., heroin, LSD, marijuana) or Schedule 2 (opium, morphine, cocaine) drugs. We classified drug users into Schedule 1 or 2 based on which type of drug the person was most dependent.

The demographic characteristics were: female gender; age at entry; race/ethnicity—white (omitted reference group); black; other race; Hispanic; marital status—married currently, other marital status (omitted reference group); educational attainment—a count variable for <high school; high school or equivalent; community college or trade school; college graduate; post graduate; advanced degree. Finally, we included a binary variable for not employed at court program entry (employed omitted group). With these predicted values for each person, we calculated the mean probability for each court. This yielded a casemix index for each court. The index values varied from 0.30 (most favorable casemix) to 0.44 (least favorable casemix).

To measure service intensity, we used factor analysis with a Varimax rotation of these variables: numbers of days in a 12-Step program (a 12-Step program involves non-medical supportive treatment); number of days in a DWI or hybrid court, and numbers of—drug tests, scheduled drug court reviews, sanctions, and incentives. DWI courts also use formal treatment, however, not all courts record this information, so as a proxy, we use days in 12-Step program. While not a formal inpatient or outpatient form of treatment; 12-Step programs, such as Alcoholics Anonymous (AA), have been shown to have similar treatment outcomes as formal treatment programs (Timko et al., 2000, Moos and Moos, 2006, Kelly et al., 2009). Incentives were offered as a carrot to induce desired behavior, including a reduction in the penalty. We used loadings from the first factor for our index of service intensity. This factor accounted for 40.7% of the variation in the measures and was positively related to each individual measure intensity—most closely related to the number of scheduled drug court reviews and least related to the number of days in a 12-Step program.

2.4. Selecting Reference Groups

We selected a reference court or courts to which we matched casemixes of individual courts. The reference court or courts was the treatment group to which we compared several alternative comparison court groups.

After obtaining casemix and service intensity indexes for each court, we plotted court casemix indexes against court service intensity indexes (Fig. 1). As required by the terms of our data agreement with the courts, individual court names have been replaced with numbers corresponding to their placement in Figure 1. Horizontal and vertical lines correspond to mean values of casemix and service intensity, which split the area into quadrants with quadrant II containing 2 courts, quadrant IV with 1, quadrant I with 2, 1 on the border—grouped in I due to its proximity to the other court in I, and quadrant III with 3 courts, respectively. Quadrant I contains courts with relatively severe casemixes, and lower service intensities; Quadrant II, relatively severe casemixes and higher service intensities; Quadrant III has courts with relatively less severe casemixes and lower service intensities; and Quadrant IV has courts with relatively less severe casemixes and greater service intensities.

We expected greater service intensity would improve court outcomes. On the other hand, a more complex casemix may require more resources to achieve the same outcome, even after matching on individual characteristics.

2.5. Propensity Score Matching

The role of the above classification scheme was to decide on which courts to match. We now describe how we implemented PSM. Rather than match on the indexes for casemix severity and service intensity, which only affected the choice of which courts to match, on a per participant basis, we matched the characteristics of participants in a comparison court or courts with characteristics of participants in a reference court or courts.⁴ The characteristics consisted of the same variables as in the logit analysis used to predict the casemix severity index for each court. Given the PSM method we used, the number of observations in the resulting matched samples were the less than the original court samples, because of missing values or because the match did not satisfy the above caliper criterion. Since the reference courts had a relatively severe casemix, the matching process tended to find matches for more severe cases in the comparison court samples.

The final result was an average treatment effect on the treated (ATT), which compares outcomes of participants in the reference court(s) with those in the comparison court samples. The ATT effect was calculated as the difference between the value for an outcome for a participant in a reference court(s) program and the corresponding value for the matched participant in the comparison court(s) program. We computed ATTs for these outcome probabilities: program completion; improved employment status; improved educational attainment; and another DWI arrest within: one year, two years, and three years following admission to the court program. Improvements in employment and educational attainment were measured using status at time of entry into the DWI court program versus status post program. To measure improvement in employment status, we ranked employment status in three categories in decreasing order: employed over 30 hours per week; employed less than 30 hours a week; and unemployed. Individuals identified as not in labor force at entry were excluded from the analysis due to the inability to improve employment. If there was any improvement in employment status, a binary variable for improved employment status was set equal to 1. Educational attainment was classified into six categories ranging from under

⁴Generally in PSM, matching is based on characteristics of the treatment group, here the reference group. However, since we wanted to compare a control group to several reference groups, we matched on characteristics of the control group.

12 years to a graduate degree. We set a binary variable for improved educational attainment equal to 1 if the person was in a higher educational attainment category at graduation than at admission to the court program. We computed re-arrest rates for any charge for 1 year and 2 year following the date of program entry for those courts for which the data appeared to be reliable.

We performed three sets of matches. The first involved comparisons of the reference court with each of the other courts. The reference court (II-B) had the most severe casemix and the highest service intensity. In a second match, we grouped courts by quadrant. The comparisons for study outcomes between II-B and court participants in court programs in each of the other 3 quadrants illustrated in Fig. 1. In a third group of comparisons, participants in both court programs in quadrant II (more severe casemix and greater service intensity) were compared with participants in the programs of courts in each of the other quadrants. The first set has the advantage of not aggregating participants from different courts within a quadrant, which have different policies and practices. However, sample sizes of participants in the first set are often much lower than in the second and third sets.

3.0. RESULTS

Overall, 18% of participants improved their employment status after the participating in the DWI court program while 39% of participants failed to complete their programs (Table 1). While the vast majority of participants had histories of alcohol use, a minority had histories of illicit drug use. Half of participants had received substance abuse treatment prior to entry into the program. Fewer than a fifth (17%) had a history of mental illness. The mean participant had a high school diploma or equivalent, but a few participants had graduate degrees. 25% were not employed at program entry.

On average, sample persons spent 362 days in DWI courts (Table 2). The median number of days was almost the same as the mean value, 364 days. However, while the mean number of days in a 12-Step program was 129, over half of the sample did not participate in a 12-Step program. There was also substantial variability in numbers of drug tests while enrolled and in use of sanctions and incentives. A few persons were in treatment after they were no longer officially enrolled in the court program. Some persons remained in the court program and were under treatment for years. The maximum values of 893 and 897 for days in court and in a 12-Step program are maximums but not outliers. Individuals with values exceeding 2.5 years in court or in a 12-Step program were eliminated from the analysis, as the maximum amount of time allowed in a DWI court is 2.5 years (e.g., maximum values of 2,177 for days in court in our original sample and 4,682 days in a 12-Step program were not acceptable values).

We performed logit analysis on whether or not the participant completed the program with and without court fixed effects (Table 3). With few exceptions, e.g., the odds ratios for race/ethnicity, adding the court fixed effects had little effect on the estimated odds ratios and associated confidence intervals. Thus, we discuss the results for the specification that excludes court fixed effects. Parameter estimates from this specification were used to construct our court-specific casemix index.

Factors associated with a higher probability of program failure included having a mental health history (odds ratio (OR)= 1.39; 95% confidence interval (CI): 1.12–1.73), having been a schedule 2 drug user (OR=1.83; 95% CI: 1.12–3.00), and having started alcohol use before age 16 (OR=1.26; 95% CI 1.04–1.52). Holding other factors constant, prior substance abuse treatment was unrelated to failure to complete the program. With the other covariates included, race/ethnicity was unrelated to program completion. (However, in the specification with court fixed effects included, being Hispanic or black was associated with a higher

probability of failing to complete the program.) Factors associated with a lower probability of program failure included being female (OR=0.79; 95% CI: 0.65–0.95), older (OR=0.98; 95% CI: 0.98–0.99 for each year of age), currently married (OR= 0.75; 95% CI: 0.59–0.96), and having a higher educational attainment (OR=0.66; 95% CI 0.59–0.73, a 34% reduction in the odds of program failure for each educational category (1–6) attained).

Before matching, there were substantial differences in characteristics of participants by court. Even after matching, some differences in characteristics remained. However, standardized differences exceeded the 10% threshold in only a minority of cases (Table 4). Matches were better for courts located in Quadrant I compared to Quadrant II (more severe casemix and lower service intensity compared courts with relatively severe casemixes with a greater service intensity) and Quadrant III—II comparisons (less severe casemix and lower service intensity compared to more severe casemix with a greater service intensity) than they were for the Quadrant IV—Quadrant II comparisons (less severe casemix and greater service intensity compared with more severe casemix with a greater service intensity). Court I-B was on the boundary of Quadrants I and III, but we placed it in Quadrant I for purposes of our analysis.

Prior to matching, 13% of participants in Courts II-A and II-B had a mental history as compared to courts with lower service intensity (Quadrant I)—Courts I-A and I-B. 20% of participants in the latter courts had a mental health history. The total sample for Courts II-B and II-A combined was 819 and for Courts I-B and I-A, the combined sample was 1,574. After matching, 16% of participants from Courts II-B and II-A had a mental health history while 14% from Courts I-B and I-A did. The standardized difference for mental health history fell from 19.1% to 3.6%, the latter being well within usual criterion for a good match, a standardized difference percentage of under 10% in absolute value. The mean percentage with mental health histories for Courts II-B and II-A increased from 13% to 16% due to the matching, which reduced the sample from these courts from 819 to 356. No match within the 0.05 caliper criteria could be found for the remaining Court II-B and II-A participants. There were also 356 persons from Courts I-B and I-A, given one-to-one matching. For the comparison of these courts, no standardized differences exceed 10%, although before matching, standardized differences above this thresholds existed for almost all the covariates. The largest difference was for percent black; before matching the standardized difference was 55.1%. After matching, the difference was –1.15%.

Even after matching, there were substantial differences in characteristics between Courts II-B, II-A, and IV-A in mental health history, age, black race, and in educational attainment at program entry. One reason for the poorer match is that there were relatively fewer observations on which to match from the Court IV-A. Not shown, in order to conserve space, are comparable data for the other matches. In general, PSM works better when there are more observations on which to match.

Table 5 shows ATT effects and associated statistical significance levels for three sets of comparisons: Court II-B (reference) and all of the other courts (comparison, Panel A); Court II-B with II-A (reference), with courts in the other quadrants (comparison, Panel B); and Court II-B with II-A (reference) combined with the courts in the other quadrants (comparison, Panel C). The difference between panels B and C is in the comparison groups; Panel C adds Court IV-A. The overall finding is that higher service intensity, even with a more severe case mix, leads to higher rates of program completion.

More specifically, compared to Court II-B, participants in Courts IV-A, III-A, I-A, and II-A were less likely to complete their programs (Panel A). There was no difference in probability of completion for Courts I-B, III-C, and III-B, and II-B (the comparison group).

This pattern suggests that on average, the more demanding, service-intensive program offered by Court II-B did not deter completion; in fact, if anything, completion rates were higher than for courts with lower requirements for completion. Court II-B had a much higher completion rate than Court II-A did—the difference in the probabilities of completion being 0.14.

Combining individual courts by quadrant (Table 5, Panel B), Court II-B had statistically significantly higher rates for completion than all of the court groups.

With Courts II-B and II-A participants as the reference group, the majority of ATTs for failed program were statistically significant at the five percent level or higher. Differences in failure probabilities were 0.14 for Courts IV-A and III-A, III-B, and III-C.

Overall, the empirical evidence for relationships between court casemix severity/service intensity was weaker for improved employment than it is for completion. The only statistically significant differences in improved employment are in Panel C. Courts II-B and II-A had greater success in improving employment status of participants than either Courts I-B and I-A and Courts III-A, III-B, III-C did. The differences in the probabilities of improved employment were fairly large in both comparisons (0.08 and 0.10).

The findings on the relationship between casemix severity/service intensity and improved educational attainment is mixed. We could not compare Courts II-B and I-A with other courts on improving educational attainment since Court II-A did not collect data on educational attainment. We found that Court II-B had more success in improving educational attainment of its participants than Court III-A, which is consistent with the view that higher service intensity leads to better educational outcomes. However, Court III-B performed even better in this dimension than Court II-B, a finding that runs counter to our hypothesis that higher service intensity, holding other factors constant, improves program outcomes.

Although the results presented thus far suggest that courts with higher service intensity tend to achieve better results in terms of program completion, we found no statistical difference in recidivism measured by rates of re-arrest for any offense with a 2-year follow-up from the date of admission to the program (Table 6). Court II-B, the more service intensive court, had a less favorable re-arrest rate than Court I-A for a 1-year follow-up.

4.0. DISCUSSION

For most outcomes, our main finding was that higher service intensity is associated with better outcomes for court participants. Although we matched on participant characteristics, we found the proportion of difficult cases a court had (higher casemix severity) was unrelated to the outcomes we studied. This goes against our hypothesis that having a more difficult to treat casemix would have an adverse effect on outcomes because the harder to treat participants would consume a larger share of court resources.

Our measure of service intensity included both duration of service and quantity of services provided per unit of time. The empirical evidence on treatment duration, not taken from court settings, generally indicates that programs of longer duration are more effective than brief interventions (Wutzke et al., 2002, Zhang et al., 2003, McKay, 2005). However, there are diminishing marginal returns to extending treatment length (Zhang et al., 2003). Another possibility is that casemix severity did differ sufficiently among courts in our analysis sample.

The results on service intensity are important in implying that resources devoted to court programs are productive in terms of achieving better outcomes at the margin. Furthermore, these results imply that creating a DWI court without thought as to the structure or design of the treatment aspect is not likely to be productive. Research on addiction interventions has shown that adjusting treatment services to accommodate an individuals' clinical assessment results in better treatment outcomes (Marlowe et al., 2009). In terms of our courts, this means that within the observed range of service intensity, -0.5 for Court IIIA to 1.0 for Court IIB, adding resources makes a difference. Future analysis should assess whether or not the difference in benefits is worth the added cost. Assessments of benefit versus cost should include a more comprehensive measure of benefit than cost savings to the criminal justice system from reduced recidivism. The calculation of benefit should also consider benefits in terms of improved productivity, both in employment settings and as household members. Courts are at a disadvantage in treating offenders with alcohol addiction as research has shown that early intervention may be more important than the intensity of the treatment (Moos and Moos, 2003). Nevertheless, a policy of increased treatment has been shown to decrease alcohol-related motor vehicle deaths (Freeborn and McManus, 2010).

This study measured mix of participants at both the level of the court for purposes of classifying courts by casemix severity and service intensity and at the individual participant level for purposes of comparing outcomes among courts. Even within quadrants, there was substantial variation in the characteristics of participants in court programs. To obtain propensity score matches, we lost a considerable number of participants. This is at least partly attributable to the heterogeneity of casemixes among court programs. In the end, courts will oppose outcome-based comparisons until they can have confidence in the casemix adjustment process.

A more important practical impediment to outcome-based comparisons is lack of adequate data. For these comparisons to serve as a basis of allocating scarce public resources, data on a number of relevant outcome measures should be collected. For example, this study was limited in assessing effects of casemix and service intensity on recidivism because the data were often so poor. Courts should be diligent in collecting recidivism data from participants, particularly in states where a central state agency does not maintain criminal court records.

This study has several strengths. First, rather than treat court programs as homogeneous entities, this study assesses heterogeneity of treatment courts as well as how differences among such courts relate to participant outcomes. Second, although the eight courts in our sample is admittedly a small number, we included a larger number of courts than in previous studies. Our courts come from two states. Further studies should attempt to include courts from a greater number of states, but past research has not crossed state lines. Third, although there is room for improvement, our study accounts for casemix differences in participants among courts.

We acknowledge several limitations. First, most of the sample came from one state. Including other state programs would substantially add to both heterogeneity of court participants, e.g., according to race/ethnicity, and to heterogeneity in program philosophy and in details of implementation. Second, due to data limitations, our study is limited to a few outcome measures. In particular, it would be important to measure the impact of court programs on future drunk driving and DWI violations and convictions. Third, the courts were selected because they were willing to participate. However, even if all courts willing to participate in our study were better than average, the fact that we assessed differences in program outcomes resulting from differences in service intensity and in client characteristics. Although the differencing approach we used eliminated some potential bias from the selection process, some selection bias may remain.

Subject to the caveats just noted, we obtained empirical support for the hypothesis that devoting more resources to treatment is productive in terms of better outcomes, but not for the hypothesis that holding other factors constant that a more severe casemix is associated with poorer program outcomes. How the courts with more severe casemixes learn to cope with the challenge of a client group more resistant to change is most certainly an important topic for further study.

Acknowledgments

This paper was funded in part by a grant from the National Institute of Health (NIH), specifically, the National Institute on Alcohol Abuse and Alcoholism (NIAAA). The grant number is 1R21AA018168-01A1. We would like to thank the courts in Michigan for sharing their data.

References

- Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat.* 2011; 10 (2):150–161. [PubMed: 20925139]
- Bouffard JA, Bouffard LA. What works (or doesn't) in a dui court? An example of expedited case processing. *J Crim Just.* 2011; 39 (4):320–328.
- Bouffard JA, Richardson KA, Franklin T. Drug courts for dwi offenders? The effectiveness of two hybrid drug courts on dwi offenders. *J Crim Just.* 2010; 38 (1):25–33.
- Eibner C, Morral AR, Pacula RL, Macdonald J. Is the drug court model exportable? The cost-effectiveness of a driving-under-the-influence court. *J Subst Abuse Treat.* 2006; 31 (1):75–85. [PubMed: 16814013]
- Freeborn BA, Mcmanus B. Substance abuse treatment and motor vehicle fatalities. *South Econ J.* 2010; 76(4):1032–1032, 1048.
- Huddleston W, Marlowe DB. Painting the current picture: A national report on drug courts and other problem-solving court programs in the united states. *Natl Drug Court Inst US.* 2011:28.
- Kelly JF, Magilli M, Stout RL. How do people recover from alcohol dependence? A systematic review of the research on mechanisms of behavior change in alcoholics anonymous. *Addict Res Theory.* 2009; 17 (3):236–259.
- Macdonald JM, Morral AR, Raymond B, Eibner C. The efficacy of the rio hondo dui court. *Evaluation Review.* 2007; 31 (1):4–23. [PubMed: 17259573]
- Marlowe DB, Festinger DS, Arabia PL, Dugosh KL, Benastutti KM, Croft JR. Adaptive interventions may optimize outcomes in drug courts: A pilot study. *Curr Psych Rep.* 2009; 11 (5):370–376.
- Mckay JR. Is there a case for extended interventions for alcohol and drug use disorders? *Addiction.* 2005; 100 (11):1594–1610. [PubMed: 16277622]
- Michigan Drug Treatment Courts. 2010 annual report and evaluation summary. Lansing; Michigan: 2010.
- Moore KA, Harrison M, Young MS, Ochshorn E. A cognitive therapy treatment program for repeat dui offenders. *J Crim Just.* 2008; 36 (6):539–545.
- Moos R, Moos B. Long-term influence of duration and intensity of treatment on previously untreated individuals with alcohol use disorders. *Addiction.* 2003; 98:325–337. [PubMed: 12603232]
- Moos R, Moos B. Participation in treatment and alcoholics anonymous: A 16-year follow-up of initially untreated individuals. *J Clin Psychol.* 2006; 62 (6):735–750. [PubMed: 16538654]
- National Highway Traffic Safety Administration. Traffic Safety Facts. U.S. Department of Transportation; 2009. 2008 traffic safety annual assessment--highlights.
- National Highway Traffic Safety Administration. Traffic Safety Facts: 2009 Data. U.S. Department of Transportation; 2011. Passenger vehicles.
- Timko C, Moos R, Finney J, Lesar M. Long-term outcomes of alcohol use disorders: Comparing untreated individuals with those in alcoholics anonymous and formal treatment. *J Stud Alcohol.* 2000; 61 (4):529–540. [PubMed: 10928723]

- Wutzke S, Conigrave K, Saunders J, Hall W. The long-term effectiveness of brief interventions for unsafe alcohol consumption: A 10-year follow-up. *Addiction*. 2002; 97:665–675. [PubMed: 12084136]
- Zhang Z, Friedman P, Gerstein D. Does retention matter? Treatment duration and improvement in drug use. *Addiction*. 2003; 98:673–684. [PubMed: 12751985]

Highlights

We assessed how casemix severity and service intensity were related to outcomes.

Outcome measures used: failure rate; education; employment; and re-arrest.

For most outcomes, higher service intensity was associated with better outcomes.

We found that a court's casemix severity was unrelated to the outcomes we studied.

Fig. 1.
Classifying Courts
Note: the two axes denote the respective means.

Table 1

Mean Values of Individual Characteristics

Variable Name	Observations	Mean	Std. Dev.
Outcomes			
Failed to complete program	2,967	0.39	0.49
Improved employment status after program	2,701	0.18	0.38
Improved educational status after program	2,566	0.11	0.32
Substance Abuse and Health History			
Mental health history	3,844	0.17	0.37
Drug history - schedule 1	3,844	0.13	0.33
Drug history - schedule 2	3,844	0.03	0.18
Drug history - alcohol	3,844	0.86	0.35
Pre-16 addiction	3,844	0.35	0.48
Any IV drug experience	3,844	0.03	0.18
Prior substance abuse treatment	3,844	0.51	0.50
Demographic Characteristics and Employment Status			
Female	3,844	0.24	0.43
Age	3,841	32.82	11.84
Married	3,844	0.17	0.37
Hispanic	3,844	0.08	0.28
Black	3,844	0.09	0.29
Other race	3,844	0.03	0.17
Educational attainment (index)	3,844	2.03	0.94
Not employed	3,844	0.25	0.44

Explanatory variables for missing values not shown.

Educational attainment varies from ___ to ___s from 1–6.

Table 2

Frequency Distribution of Program Variables

	Days in DWI court	Days in 12 step program	No. of scheduled DWI court reviews	No. of drug tests	No. of sanctions	No. of incentives
Percentile						
1	13	0	0	0	0	0
25	200	0	4	28	0	0
50	364	0	9	105	1	1
75	505	189	19	224	3	3
100	893	897	116	828	30	23
Mean	362.40	128.78	13.35	142.95	1.84	2.02
Std. Dev.	197.84	208.05	13.26	138.77	2.42	2.83
Observations	3,844	3,844	3,839	3,838	3,814	3,814

Table 3

Failure to Complete Drug Court Program (Logit Analysis)

Variables	Fixed Effects?	
	No	Yes
Substance Abuse and Health History		
Mental health history	1.39 (1.12–1.73)	1.43 (1.14–1.79)
Drug history - schedule 1	1.19 (0.87–1.63)	1.22 (0.88–1.69)
Drug history - schedule 2	1.83 (1.12–3.00)	1.97 (1.19–3.25)
Drug history - alcohol	0.75 (0.52–1.08)	0.73 (0.50–1.06)
Pre-16 addiction	1.26 (1.04–1.52)	1.36 (1.12–1.65)
Any IV drug experience	0.94 (0.60–1.47)	0.88 (0.56–1.39)
Prior substance abuse treatment	1.10 (0.93–1.30)	1.10 (0.92–1.30)
Demographic Characteristics and Employment Status		
Female	0.79 (0.65–0.95)	0.76 (0.62–0.92)
Age	0.98 (0.98–0.99)	0.98 (0.98–0.99)
Married	0.75 (0.59–0.96)	0.74 (0.58–0.95)
Hispanic	1.15 (0.84–1.58)	1.69 (1.20–2.37)
Black	1.16 (0.86–1.55)	1.59 (1.19–2.24)
Other race	1.01 (0.61–1.68)	1.18 (0.70–1.97)
Unemployed	1.88 (1.57–2.26)	1.97 (1.64–2.38)
Educational attainment (index)	0.66 (0.59–0.73)	0.65 (0.58–0.73)
Observations	2,967	2,967

95% confidence intervals in parentheses.

Explanatory variables for missing values not shown.

Standardized Differences: Pre- and Post-Matching

Table 4

Courts	Pre-Match				Post-Match				Pre-Match				Post-Match				
	I-B vs. I-A	II-B vs. II-A	Std. Diff. (%)	I-B--I-A	II-B--II-A	Std. Diff. (%)	IV-A	Std. Diff. (%)	IV-A	II-B--II-A	Std. Diff. (%)	III-A--III-B-- III-C	IIB--IIA	Std. Diff. (%)	III-A--III-B-- III-C	II-B--II-A	Std. Diff. (%)
Substance Abuse and Health History																	
Mental health history	0.20 (0.40)	0.13 (0.33)	19.09	0.14 (0.35)	0.16 (0.36)	-3.56	0.28 (0.45)	0.13 (0.33)	38.01 (0.44)	0.26 (0.44)	0.22 (0.42)	0.11 (0.31)	0.13 (0.33)	-6.25	0.14 (0.35)	0.14 (0.35)	0.11
Drug history - schedule 1	0.19 (0.39)	0.084 (0.28)	31.15	0.12 (0.35)	0.13 (0.34)	-2.93	0.064 (0.25)	0.084 (0.28)	-7.74 (0.25)	0.069 (0.25)	0.11 (0.32)	0.083 (0.28)	0.084 (0.28)	-0.60	0.10 (0.31)	0.11 (0.32)	-4.54
Drug history - schedule 2	0.038 (0.19)	0.034 (0.18)	2.16	0.028 (0.17)	0.033 (0.18)	-2.88	0.030 (0.17)	0.034 (0.18)	-2.28 (0.18)	0.033 (0.18)	0.041 (0.20)	0.023 (0.15)	0.034 (0.18)	-6.64	0.028 (0.17)	0.034 (0.18)	-3.24
Drug history - alcohol	0.85 (0.36)	0.85 (0.36)	0.00	0.83 (0.38)	0.82 (0.39)	3.28	0.90 (0.30)	0.85 (0.36)	15.09 (0.29)	0.91 (0.29)	0.87 (0.34)	0.85 (0.36)	0.85 (0.36)	0.00	0.81 (0.39)	0.82 (0.39)	-0.58
Pre-16 addiction	0.32 (0.47)	0.50 (0.50)	-37.08	0.44 (0.50)	0.43 (0.50)	2.02	0.45 (0.50)	0.50 (0.50)	-10.19 (0.50)	0.47 (0.50)	0.60 (0.49)	0.24 (0.43)	0.50 (0.50)	-55.10	0.38 (0.49)	0.43 (0.50)	-11.07
Any IV drug experience	0.038 (0.19)	0.021 (0.14)	10.40	0.023 (0.15)	0.023 (0.15)	0.00	0.080 (0.27)	0.021 (0.14)	27.70 (0.21)	0.045 (0.21)	0.041 (0.20)	0.024 (0.15)	0.021 (0.14)	2.24	0.025 (0.16)	0.025 (0.16)	0.00
Prior substance abuse treatment	0.41 (0.49)	0.59 (0.49)	-35.32	0.56 (0.50)	0.60 (0.49)	-7.45	0.73 (0.44)	0.59 (0.49)	30.64 (0.47)	0.67 (0.47)	0.67 (0.47)	0.50 (0.50)	0.59 (0.49)	-17.64	0.63 (0.48)	0.62 (0.49)	1.73
Demographic Characteristics and Employment Status																	
Female	0.25 (0.44)	0.19 (0.39)	15.20	0.21 (0.41)	0.20 (0.40)	1.25	0.24 (0.43)	0.19 (0.39)	12.96 (0.43)	0.24 (0.43)	0.20 (0.40)	0.27 (0.45)	0.19 (0.39)	19.76	0.22 (0.41)	0.22 (0.41)	1.05
Age	31.45 (12.37)	33.60 (10.70)	-18.58	31.73 (12.01)	32.55 (10.56)	-7.24	37.10 (11.16)	33.60 (10.70)	32.03 (11.06)	35.67 (11.06)	32.25 (10.88)	32.70 (11.82)	33.60 (10.70)	-7.95	32.83 (11.36)	32.64 (10.64)	1.76
Hispanic	0.070 (0.26)	0.22 (0.42)	-44.31	0.14 (0.34)	0.14 (0.35)	-2.20	0.0028 (0.053)	0.22 (0.42)	-74.13 (0.64)	0.0041 (0.064)	0.0081 (0.090)	0.022 (0.15)	0.22 (0.42)	-64.27	0.050 (0.22)	0.056 (0.23)	-2.67
Black	0.052 (0.22)	0.24 (0.43)	-55.02	0.13 (0.33)	0.13 (0.34)	-1.15	0 (0)	0.24 (0.43)	-78.93 (0)	0 (0)	0.27 (0.45)	0.071 (0.26)	0.24 (0.43)	-47.56	0.15 (0.36)	0.14 (0.35)	2.38
Other	0.024 (0.15)	0.017 (0.13)	4.57	0.020 (0.14)	0.020 (0.14)	0.00	0.061 (0.24)	0.017 (0.13)	22.72 (0.20)	0.041 (0.20)	0.033 (0.18)	0.035 (0.18)	0.017 (0.13)	11.40	0.036 (0.19)	0.025 (0.16)	6.50
Married	0.15 (0.36)	0.21 (0.41)	-15.55	0.15 (0.36)	0.17 (0.38)	-6.14	0.22 (0.41)	0.21 (0.41)	2.16 (0.41)	0.22 (0.42)	0.15 (0.36)	0.14 (0.35)	0.21 (0.41)	-18.36	0.17 (0.37)	0.15 (0.36)	3.95
Unemployed	0.29 (0.45)	0.23 (0.42)	13.08	0.20 (0.40)	0.23 (0.42)	-7.39	0.24 (0.43)	0.23 (0.42)	2.98 (0.42)	0.23 (0.42)	0.29 (0.45)	0.23 (0.42)	0.23 (0.42)	0.40	0.23 (0.42)	0.21 (0.41)	4.05
Educational attainment (index)	1.94 (0.85)	1.89 (1.085)	5.12	1.94 (0.84)	1.90 (1.11)	3.59	2.26 (0.97)	1.89 (1.085)	35.98 (0.96)	2.25 (0.96)	2.06 (0.95)	2.20 (0.91)	1.89 (1.085)	30.95	2.03 (0.77)	1.97 (1.12)	5.54
N	1,574	819		396	396		361	819		246	246	1,090	819		357	357	

Explanatory variables for missing values not shown: mental health history, drug history – schedule 1, drug history -schedule 2, drug history - alcohol, pre-16 addiction, any IV drug experience, prior substance abuse treatment, female, current age, hispanic, black, other race, educational attainment, unemployed.

Table 5

Program Completion, Improved Employment, and Improved Educational attainment (ATT Analysis)

PANEL A									
Court	Reference (II-B)	Comparison [*]	Difference ^{**}	T-Test	Reference (II-B)	Comparison	Difference	T-Test	T-Test
	Failed Program				Improved Employment				
I-B	0.28	0.38	-0.092	-1.58	0.22	0.14	0.08	1.66	1.47
III-C	0.11	0.28	-0.17	-2.22	0.27	0.15	0.12	1.44	-0.83
IV-A	0.23	0.47	-0.25	-4.38	0.22	0.23	-0.0081	-0.15	0.23
III-B	0.26	0.38	-0.12	-1.71	0.17	0.29	-0.12	-1.93	-3.26
III-A	0.27	0.53	-0.25	-4.75	0.23	0.15	0.074	1.62	1.96
I-A	0.29	0.51	-0.22	-4.48	0.20	0.16	0.045	1.10	-0.53
II-A	0.28	0.29	-0.011	-0.16	0.17	0.28	-0.11	-1.83	-
PANEL B									
	Reference (II-B/II-A)	Comparison	Difference	T-Test	Reference (II-B/II-A)	Comparison	Difference	T-Test	T-Test
	Failed Program				Improved Employment				
I-B--I-A	0.27	0.45	-0.18	-3.72	0.21	0.15	0.055	1.36	-1.00
III-A--III-B--III-C	0.26	0.51	-0.25	-4.81	0.22	0.16	0.057	1.28	-0.87
PANEL C									
	Reference (II-B/II-A)	Comparison	Difference	T-Test	Reference (II-B/II-A)	Comparison	Difference	T-Test	T-Test
	Failed Program				Improved Employment				
I-B--I-A	0.34	0.39	-0.06	-1.62	0.22	0.15	0.07	2.41	
IV-A	0.31	0.42	-0.11	-2.44	0.27	0.22	0.06	1.41	
III-A--III-B--III-C	0.32	0.42	-0.10	-2.80	0.23	0.20	0.03	0.88	

Bold indicates significance at 5% level or higher

* Comparison is for court indicated in the left most column.

** Difference is between the reference and comparison courts.

Table 6

Recidivism (ATT Analysis)

Court	Reference	Comparison *	Difference **	T-Test	Reference	Comparison	Difference	T-Test	Reference	Comparison	Difference	T-Test
Rearrests												
1-Year Rearrests												
IV-A	0.11	0.12	-0.0057	-0.17	0.080	0.068	0.011	0.41	0.11	0.11	0.0057	0.17
I-A	0.11	0.091	0.024	0.88	0.083	0.051	0.031	1.42	0.11	0.075	0.035	1.38
2-Year Rearrests												

Bold indicates significance at 5% level or higher

* Comparison is for court indicated in the left most column.

** Difference is between the reference and comparison courts.