

**A SYSTEMATIC REVIEW OF DWI COURT  
PROGRAM EVALUATIONS**

**By Douglas B. Marlowe, J.D., Ph.D.,  
David S. Festinger, Ph.D., Patricia L. Arabia, M.S., Jason  
R. Croft, B.A., Nicholas S. Patapis, Psy.D., M.A.C.J.,  
and Karen L. Dugosh, Ph.D.**

**Treatment Research Institute at the  
University of Pennsylvania**

*A systematic literature review was conducted of published and unpublished DWI Court program evaluations released through April 30, 2007. [See Addendum for a subsequent evaluation released in October 2007]. Each evaluation report was scored for methodological rigor by at least two trained, independent raters according to established scientific criteria. One evaluation exceeded 80% of recommended criteria (deemed methodologically “good”) and an additional four evaluations exceeded 65% of recommended criteria (deemed “marginally acceptable”). Many of the evaluations had serious methodological shortcomings, including reporting outcomes only for graduates, failing to account for participant dropout, employing inadequate statistical techniques, and evaluating potentially immature programs. Although the results hint at emerging evidence potentially favoring the effects of DWI Courts, it is not possible to reach scientifically defensible conclusions about the effects of DWI Courts due to the inadequate state of the evaluation literature. It is hoped the methodological criteria outlined in this review article will influence future DWI Court program evaluations and assist practitioners and policymakers to become competent and effective consumers of evaluation findings.*

*This project was supported by a generous grant from The Century Council. The conclusions drawn are those of the authors and do not necessarily reflect the views of*

*The Century Council or the National Drug Court Institute. The authors gratefully acknowledge Emily James and Michele Pich for their assistance in collecting and rating evaluation reports.*

*Douglas B. Marlowe, J.D., Ph.D., is a Senior Scientist at the Treatment Research Institute, the Chief of Science, Policy & Law for the National Association of Drug Court Professionals, and an Adjunct Associate Professor of Psychiatry at the University of Pennsylvania School of Medicine. His research focuses on examining the role of coercion in drug abuse treatment, the effects of drug courts and other diversion programs for drug abusing offenders, and behavioral treatments for drug abusers and offenders.*

*David S. Festinger, Ph.D., is the Director of the Section on Law & Ethics Research at the Treatment Research Institute and an Adjunct Assistant Professor of Psychiatry at the University of Pennsylvania School of Medicine. His research focuses on evaluating the clinical and ethical impacts of coercive interventions for substance-abusing criminal offenders.*

*Patricia L. Arabia, M.S., is the Section Coordinator for the Section on Law & Ethics Research at the Treatment Research Institute. She is responsible for the operational management of all criminal justice and related projects, including data collection and analyses, supervision of research assistants, and monitoring of study activities.*

*Jason R. Croft, B.A., is a Research Coordinator for the Section on Law & Ethics Research at the Treatment Research Institute. He oversees on-site activities for several research projects involving drug courts and research ethics.*

*Nicholas S. Patapis, Psy.D., M.A.C.J., is a Scientist in the Section on Law & Ethics Research at the Treatment Research Institute.*

*Karen L. Dugosh, Ph.D., is a quantitative psychologist who provides methodological and statistical expertise for scientists at the Treatment Research Institute. She is skilled in cross-sectional and longitudinal data analyses, predictive modeling, and psychometric analyses.*

Direct all correspondence to: Douglas B. Marlowe, J.D., Ph.D.  
Treatment Research Institute at the University of Pennsylvania,  
600 Public Ledger Bldg., 150 S. Independence Mall West,  
Philadelphia, PA 19106-3475. (215) 399-0980. (215) 399-0987  
(fax). [dmарlowe@tresearch.org](mailto:dmарlowe@tresearch.org)

## ARTICLE SUMMARIES

### **SYSTEMATIC REVIEW OF DWI COURTS**

[1] A systematic literature review was conducted of published and unpublished evaluations of DWI Courts released through April 30, 2007.

### **RECENT EVALUATION OF DWI COURTS**

[3] A newer evaluation released in October 2007 lends additional promising support for DWI Courts.

### **EFFECTS OF DWI COURTS**

[2] Many evaluations had serious methodological shortcomings. Although results hint at emerging evidence favoring DWI Courts, it is not possible to reach scientifically defensible conclusions due to the inadequate state of the evaluation literature.

## INTRODUCTION

Approximately 40% of traffic accidents and fatalities in the U.S. are alcohol related (Greenfield, 1998; NHTSA, 1998). A partially overlapping 20% involve abuse of illicit drugs alone or in combination with alcohol (Compton & Anderson, 1985; Marzuk et al., 1990; NIDA, 2005; Simpson et al., 2006). Although the majority of individuals arrested for driving while impaired (DWI)<sup>1</sup> do not go on to repeat the offense, between 20% and 35% will become recidivist DWI offenders (e.g., Cornish & Marlowe, 2003; Timken, 2002).

A number of policy initiatives have been aimed at reducing DWI conduct in the general population. These include increasing the legal drinking age, lowering the presumptive BAC level for impaired driving, and establishing random sobriety checkpoints. Such measures have been associated with significant reductions of approximately 7% to 15% in traffic accidents and fatalities (Shults et al., 2001; Wagenaar et al., 1995). The positive effects of these policies are generally attributed to deterring first-time DWI offenders as opposed to altering the conduct of individuals already engaged in recidivist DWI behaviors (e.g., Popkin & Wells-Parker, 1994).

Among individuals who have been arrested for DWI, a range of punitive and incapacitating sanctions may be applied. These include driver's license suspension or revocation, jail terms, fines, mandatory vehicle sales, and ignition interlock requirements. Evidence suggests such measures can elicit moderate reductions in DWI recidivism of approximately 5% to 10%; however, the effects often wane

---

<sup>1</sup> The term driving while impaired (DWI) is used generically in this article to encompass comparable offense terminology, including driving under the influence (DUI) and driving while intoxicated (DWI).

after the constraints are removed (Timken, 2002; Wagenaar & Maldonado-Molina, 2007; Wagenaar et al., 1995). Moreover, it appears such sanctions may be least effective for substance dependent individuals or those with other high-risk factors for DWI recidivism, including social isolation, poor educational or employment skills, serious criminal histories, or co-morbid psychiatric conditions (e.g., Popkin & Wells-Parker, 1994; Yu, 2000).

Approximately 30% to 50% of DWI offenders satisfy official diagnostic criteria for substance abuse or dependence (e.g., Timken, 2002). For these individuals, an integrated strategy that combines license restriction, sanctions and substance abuse treatment elicits the best results (DeYoung, 1997). A comprehensive meta-analysis concluded that substance abuse treatment or remedial education contributed an additional 8% to 9% reduction in DWI recidivism for problem drinkers over punitive approaches (Wells-Parker et al., 1995). Unfortunately, compliance with substance abuse treatment is often unacceptably poor as evidenced by high rates of premature dropout (e.g., Ball et al., 2006; Festinger et al., 2002; Simpson et al., 1997; Stark, 1992). Moreover, many DWI offenders fail to comply with other restrictive conditions of supervision, such as failing to install ignition interlocks and continuing to drive on a suspended or revoked license (e.g., McCartt, et al., 2002; Robertson et al., 2007; Timken, 2002).

DWI Courts were created to improve recidivist DWI offenders' compliance with substance abuse treatment and other supervisory conditions (Freeman-Wilson & Huddleston, 1999). Modeled after Drug Courts, DWI Courts require participants to attend on-going status hearings in court, complete an intensive regimen of substance abuse treatment along with indicated adjunctive services, and undergo random or continuous biological testing for substance ingestion (NDCI, 2006). Participants receive negative sanctions for program infractions and positive rewards for achievements

that steadily increase in magnitude over successive instances. The vast majority of DWI Courts are post-adjudication programs. Many require participants to serve some portion of an incarcerative sentence, with the remainder of detention being suspended pending completion of treatment. Failure to successfully graduate from the DWI Court typically results in a return to custody to complete the full sentence. As of December 31, 2007, there were 110 separately designated DWI Courts and an additional 286 hybrid DWI/Drug Courts in the U.S. (Huddleston, Marlowe & Casebolt, 2008).

[1] The current project involved a systematic literature review of DWI Court program evaluations released through April 30, 2007. [See Addendum for a subsequent evaluation released in October 2007]. To avoid a “publication bias” resulting from the fact that negative findings are less likely to make their way into the peer-reviewed literature, both published and unpublished evaluation reports were solicited. All evaluation reports were scored according to established scientific review criteria by at least two independent raters. To our knowledge, this is the first effort to systematically evaluate the state of research on DWI Court programs using standardized criteria for methodological rigor.

## **METHODS**

### **Search Strategy**

Published and unpublished DWI Court evaluation reports were collected through April 30, 2007, the official cut-off date for this systematic review. [See Addendum for a subsequent evaluation released in October 2007]. Unpublished reports were solicited from statewide problem-solving court coordinators and other primary points of contact (PPCs) in every state and territory in the U.S. The National Drug

Court Institute (NDCI) maintains a list of PPCs who are primarily responsible for tracking statewide problem-solving court activity in their respective jurisdictions. These individuals are typically employees of the state Supreme Court, administrative office of the courts, governor's office or single state agency for substance abuse services. In addition, many are officers of their state or regional drug court associations or representatives of the Congress of State Drug Court Associations.

The PPCs and statewide problem-solving court coordinators were contacted by phone, e-mail and in person at the annual meeting of the National Network of State and Territorial Drug Court Coordinators to solicit any and all evaluation reports that were available on DWI Courts in their jurisdictions. At least three follow-up reminders were sent to each individual who did not respond to a prior contact. Out of 53 states and territories, representatives of 29 (55%) responded to the solicitations. Of those, 17 provided at least one evaluation report and 12 indicated their jurisdiction either had no DWI Court, no evaluation had been completed, or the evaluation report was not yet available.

A literature search was also conducted of published studies on relevant electronic databases, including PubMed, Medline, PsychINFO and the Computer Retrieval of Information on Scientific Projects (CRISP). The CRISP database describes federally funded biomedical research projects conducted at universities, hospitals and other research institutions. An exhaustive list of logically derived search terms was entered into each database. The search terms and number of "hits" returned for each term are presented in Table 1. Abstracts of all citations returned from the searches were reviewed to determine their face validity by two doctoral-level scientists (i.e., whether they appeared to be reporting on the evaluation of a DWI Court program).



Table 1. Search Terms for Published Electronic Databases

<b>Term</b>	<b>Hits</b>	<b>Term</b>	<b>Hits</b>
DUI	226	“Drug court” + DWI	0
“DUI Court”	18	“Drug court” + DUI	1
DUI + Driving	157	“Impaired Driving”	0
DUI + Court	0	“Program Evaluation” + DWI	5
DWI*	1268	DUI + Traffic	68
“Driving while Intoxicated”	173	DWI + Traffic	79
DWI + Driving	160	“Drunk Driving”	255
DWI + Court	19	“Drunk Driving” + Court	11
“Drunk Driving Court”	20	“Drunk Driving” + Diversion	2
“Driving Under the influence”	347	“Drunk Driving” + “Treatment Court”	18
“Driving Under the influence” + Court	22	“Specialty Court”	0
“Drug Court”	35	“Problem Solving Court”	0
“Drug Court” + Traffic	0	“Treatment Court”	18

## **Pre-Screening**

A total of 41 published and unpublished evaluations were identified from the above sources. These, in turn, were subjected to a pre-screening process to confirm that they were reporting outcomes from a DWI Court program evaluation. For example, several reports were of process evaluations and did not present client-level outcomes, such as alcohol use or recidivism. Others appeared to be reporting on a DWI Court, but further examination revealed they were actually reporting on a DWI treatment program or DWI probation track. Finally, several evaluations were of hybrid DWI/Drug Court programs and did not report the results separately for DWI offenders. Therefore, it was not possible to analyze the effects of the programs for DWI offenders.

Each report was independently reviewed by two trained raters to confirm that all of the following criteria for inclusion were met:

1. The participants must have been charged with a DWI offense.
2. The program must have involved a separately identified court docket or calendar as opposed to being administered by probation or a treatment program.
3. At least one client-level outcome must have been reported (e.g., criminal recidivism or alcohol use).
4. If the program was a hybrid DWI/Drug Court, outcomes must have been analyzed and reported separately for DWI offenders.

Prior to conducting the pre-screening, each rater completed a full-day didactic training on standardized procedures for coding critical aspects of evaluation studies according to established scientific criteria. Subsequently, the raters independently co-rated a minimum of six practice reports followed by discrepancy reviews. In most instances,

there was 100% exact agreement between the raters. In those instances when there were coding discrepancies, the raters met together with the principal investigators to resolve the discrepancies and develop explicit decision rules for handling similar issues in the future.

A total of 27 evaluations were excluded because they did not report client-level outcomes (n = 11), did not involve separately identified court dockets or calendars (n = 21) or did not report outcomes separately for DWI offenders (n = 9) (some reports were excluded for multiple reasons). Fourteen evaluations were retained for substantive review.

### **Methodological Quality Score (MQS)**

The remaining 14 evaluations were scored by at least two independent raters for methodological rigor according to standardized review criteria. A Methodological Quality Score (MQS) was assigned to each evaluation pursuant to a scoring system adapted from the *Mesa Grande Coding System for Methodological Quality* (Miller & Wilbourne, 2002). The scoring criteria for the MQS are presented in Table 2.

Table 2. Criteria for Methodological Quality Score (MQS)\*

Study Design	1 = Single group, post-test only (e.g., comparison to national data) 2 = Single group, pre-to-post change 3 = Non-randomized comparison sample (e.g., drawn from a neighboring county or before the DWI Court was started) 4 = Quasi-experimental; i.e., drawn from the same population at the same time (e.g., waitlist control) 5 = Randomized
Replicability	0 = Procedures not described in sufficient detail to permit replication 1 = Procedures described in sufficient detail re. inclusion criteria, comparison sample, assessments, interventions and statistical analyses
Baseline Measures	0 = No client-level baseline measures reported 1 = Basic demographic information reported (e.g., age, gender, education) 2 = Baseline measures reported that are predictive of outcomes (e.g., severity of alcohol problem, DWI criminal history, prior treatment history)
Quality Control	0 = Program not standardized or described 1 = Interventions standardized by manual, procedures, training, etc.

Table 2 continues . . .

Follow-up Interval	0 = Follow-ups conducted during treatment only 1 = Follow-ups conducted < 6 months post-discharge 2 = Follow-ups conducted 6 to 11 months post-discharge 3 = Follow-ups conducted $\geq$ 12 months post-discharge
Dosage	0 = No discussion of the dosage or % of services received 1 = One of the following was reported: dosage of treatment services, court/criminal justice contacts or biological testing received 2 = Two of the following were reported: dosage of treatment services, court/criminal justice contacts or biological testing received 3 = Dosage of treatment services, court/criminal justice contacts and biological testing all reported
Collaterals	0 = No collateral verification of clients' self-report (e.g., family members, employers) 1 = Collaterals interviewed
Objective Verification	0 = No objective verification of clients' self-report 1 = Verification of client reports (e.g., criminal records, urine screens, pay stubs)
Dropout/attrition	0 = Dropouts not accounted for 1 = Dropouts enumerated and attrition statistically accounted for

*Table 2 continues . . .*

Statistical Power	0 = Inadequate statistical power due to small sample size or attrition 1 = Adequate statistical power with adequate sample size
Analyses	0 = No statistical analyses or clearly inappropriate analyses 1 = Appropriate statistical analyses
Generalizability	0 = Evaluation conducted at single site 1 = Parallel evaluations conducted at two or more sites
Follow-up Rate	0 = < 70% follow-ups completed 1 = 70 to 84% follow-ups completed 2 = 85 to 100% follow-ups completed

*\*Adapted from Miller & Wibourne (2002); Vaughn & Howard (2004)*

The *Mesa Grande Coding System* was selected for several reasons. First, it is the only coding system specifically developed for evaluations of substance abuse treatment interventions (e.g., Becker & Curry, 2008; Miller et al., 1995). Second, it has been frequently used in reviews of various types of substance abuse treatments (Miller & Wilbourne, 2002; Vaughn & Howard, 2004) and therefore provides a basis for comparing the quality of DWI Court evaluations against those of other substance abuse programs. Third, unlike coding systems such as CONSORT that were developed for tightly controlled, experimental studies (Moher et al., 2001; Moja et al., 2005), the MQS employs more liberal scoring criteria that can be used for evaluations in “real-world” treatment settings. For example, it applies partial credit for non-randomized designs and does not require strict adherence to treatment manuals or therapist-competency measures.

The *Mesa Grande* system does, however, require scientifically defensible evaluation designs that permit inferences of causality to be reached about the effects of the programs. Some DWI Courts may lack sufficient resources or scientific expertise to pass muster, even under this more liberal scoring system; however, the alternative of lending credence to unreliable findings is not acceptable for a systematic literature review. Importantly, it should be recognized that the MQS assesses the quality of the evaluation designs, and not the quality of the DWI Courts themselves. The MQS criteria generally relate to research procedures and statistical analyses and do not address matters of professional competence, training or team functioning.

The MQS ranges from 0 to 23 with higher scores reflecting greater methodological rigor. A score satisfying at least 80% of recommended criteria (i.e.,  $MQS \geq 19$  out of 23) was considered to be “good” and a score satisfying at least 65% of recommended criteria ( $MQS \geq 15$ ) was considered to be “marginally acceptable” (cf., Miller & Wilbourne, 2002).

A relatively liberal cut-off score of 65% was set for marginally acceptable evaluations because most DWI Court evaluations are conducted in real-world programs by local evaluators and not in scientifically controlled research settings. Setting more stringent criteria could have the effect of excluding evaluations that provide useful and practical information about how these programs perform in day-to-day practice.

As with the pre-screening process, the raters were required to complete a didactic training on standardized coding procedures and independently evaluated at least six practice reports. No rater participated in the project until he or she attained  $\geq .80$  inter-rater reliability (IRR) with other raters on anchoring protocols. All scoring discrepancies were resolved as a group with the principal investigators and the agreed-upon scores were used in substantive data analyses.

### **Program Maturity Index (PMI)**

Each DWI Court was also assigned a Program Maturity Index (PMI) reflecting the number of years it had been in operation prior to the initiation of the evaluation. Generally speaking, data collected during the first year of operations should be used to inform programmatic modifications, and should ordinarily be included in a process analysis as opposed to an outcome analysis (e.g., Heck, 2006; Rempel, 2007). Ideally, DWI Courts should be given ample time to pilot-test their operations and implement indicated modifications before outcome analyses are conducted. The PMI was not included as part of the MQS because it does not relate to the evaluation procedures, but rather to the experience and maturity of the program.

The PMI ranged from 0 to 3 with higher scores indicating longer-standing programs. The PMI scores were assigned according to the following criteria:



- 0 = program was  $< 1$  year old
- 1 = program was  $\geq 1$  year old and  $< 2$  years old
- 2 = program was  $\geq 2$  years old and  $< 3$  years old
- 3 = program was  $\geq 3$  years old

Importantly, some evaluations were conducted over an extended period of time and reported outcomes for participants who entered during the first year of operations as well as during subsequent years. If it was not possible to disentangle the results of the first year of operations from those of subsequent years, the evaluation received a PMI of 0 (i.e.,  $< 1$  year).

## RESULTS

The proportion of evaluations satisfying various methodological criteria is presented in Table 3. One-half (50%) of the evaluations employed non-randomized comparison samples, such as DWI offenders who were arrested in a neighboring county or prior to the establishment of the DWI Court program. Twenty-nine percent of the evaluations were single-group studies that compared outcomes to national data, and 21% were randomized experiments.

Table 3. Proportion of Evaluations Satisfying Methodological Criteria: n (%)

Study Design	Single group, post-test	4(29%)
	Single group, pre-to-post change	0(0%)
	Non-randomized comparison sample	7(50%)
	Quasi-experimental	0(0%)
	Randomized	3(21%)
Replicability	Procedures not described in sufficient detail	5(36%)
	Procedures described in sufficient detail	9(64%)
Baseline Measures	Client-level variables not reported	5(36%)
	Basic demographics reported	1(7%)
	Predictors of outcomes reported	8(57%)
Quality Control	Interventions not described or standardized	3(21%)
	Interventions described and standardized	11(79%)
Follow-up Interval	Follow-ups during treatment only	6(43%)
	Follow-ups < 6 months post-discharge	1(7%)
	Follow-ups 6 to 11 months post-discharge	2(14%)
	Follow-ups ≥ 12 months post-discharge	5(36%)

Table 3 continues . . .

Dosage	No dosage of services reported	11(79%)
	Dosage of one service reported	1(7%)
	Dosages of two services reported	0(0%)
	Dosages of three services reported	2(14%)
Collaterals	No collateral verification of client reports	14(100%)
	Collaterals interviewed	0(0%)
Verification	No objective verification of client reports	4(29%)
	Verification of client reports	10(71%)
Dropout/attrition	Dropouts not accounted for	5(36%)
	Dropouts accounted for	9(64%)
Statistical Power	Inadequate statistical power	8(57%)
	Adequate statistical power	6(43%)
Analyses	No statistical analyses or inappropriate analyses	9(64%)
	Appropriate statistical analyses	5(36%)
Generalizability	Single site	13(93%)
	Two or more sites	1(7%)
Follow-up Rate	< 70% follow-up rate	6(43%)
	70% to 84% follow-up rate	0(0%)
	85% to 100% follow-up rate	8(57%)

Roughly two-thirds (64%) of the evaluations described the research methods in sufficient detail to permit replication by other investigators. Over three-quarters (79%) of the DWI Court programs followed a standardized regimen that was sufficiently described in the evaluation report to permit the reader to understand the type of program that was being assessed.

Nearly three-quarters of the evaluations (71%) reported on objectively verifiable outcome measures, such as urine results or graduation rates; however, none collected information from collateral persons, such as family members or employers. Unfortunately, nearly two-thirds of the evaluations failed to properly account for participant dropout (64%) or used inappropriate or no statistical analyses (64%). The evaluations were about evenly split in terms of whether they had a large enough sample size for statistical power (43%), achieved a minimally adequate follow-up rate of at least 70 percent (57%) and measured outcomes over a period of at least six months post-discharge (50%).

A large proportion (79%) of the evaluations failed to report any information on the dosages of services that were actually received by participants, such as the number of counseling sessions or status hearings that were attended (as opposed to what was planned or scheduled). Only 14% of the evaluations reported dosage information on several key services for a DWI Court program, including counseling sessions, court hearings and biological tests for substance use. As a result, it was not possible in most instances to determine which components of the programs, if any, might have contributed to effective outcomes or how well the programs were implemented in practice.

Figure 1 depicts the distribution of MQS scores for the 14 evaluations. One evaluation exceeded 80% of recommended methodological criteria and an additional four evaluations exceeded 65% of recommended criteria.

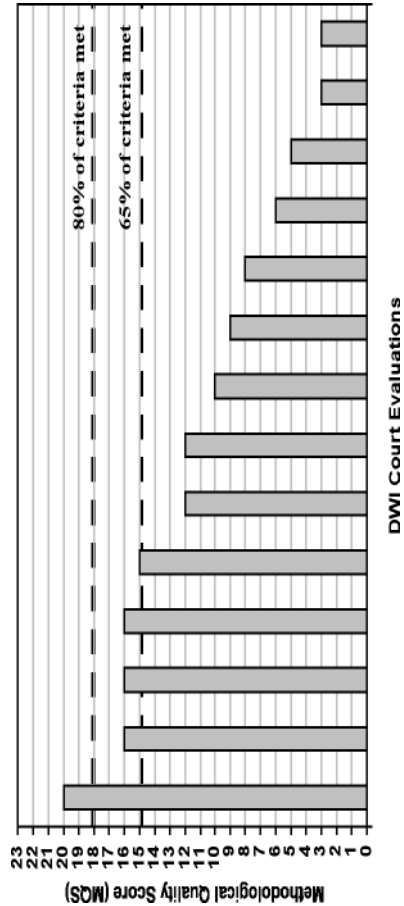


Figure 1. Distribution of Methodological Quality Scores (MQS) for DWI Court Program Evaluations Released Through April 30, 2007

The five evaluations satisfying at least 65% of recommended criteria are summarized in Table 4 and described below. Of these, three had PMI scores of 0 indicating they evaluated the programs, at least in part, during the first year of operations. The remaining two programs had been in operation for at least two years prior to initiating the evaluation.

Table 4. Summary of DWI Court Program Evaluations with Good to Marginally Acceptable MQS Scores

Citation	MQS (0 – 23)	PMI (0 – 3)	Primary Outcome(s)	Follow-up Interval	Comparison Sample	Summary of Results
Jones (2005)	20	0	Convictions for alcohol-related traffic offenses	2 years post-entry	Random assignment to DWI probation	Intent-to-treat sample: n.s.  Completers: 3.6% vs. 6.4% $p < .05$ , ES = .13 (small)
MacDon- ad et al. (2007)	16	0	Arrests for alcohol- related offenses	2 years post-entry	Random assignment to DWI probation	n.s.
			Self-reported DWI events			n.s.
			Self-reported alcohol abuse			n.s.
						<i>Table 4 continues . . .</i>





			Arrests for DWI offenses				Graduates: 5% vs. 19% <i>P</i> < .05, ES = .45 (medium)
Brecken- ridge et al. (2000)	15	2	Convictions for traffic offenses	24 months post-arrest	Random assignment to adjudication as-usual		n.s.
			Convictions for alcohol, drug or other serious offenses				

Notes: MQS = Methodological Quality Score. PMI = Program Maturity Index. n.s. = not statistically significant. ES = effect size expressed as *h* for proportional data.

\* Reflects an intensive court supervised probation program.

### **Good to Marginally Acceptable Evaluations of Immature Programs**

Three evaluations employed good to marginally acceptable research methodology, but involved programs that had been in operation for only a short period of time. As a result, the implications of the findings for the efficacy of those DWI Courts remain somewhat questionable.

#### *Maricopa County DWI Court.*

The evaluation receiving the highest MQS (20 out of 23; 86%) was an experimental study of the Maricopa County (Arizona) DWI Court (Jones, 2005). Individuals convicted of a felony DWI offense were randomly assigned either to the DWI Court ( $n = 387$ ) or to the county's standard probation program ( $n = 397$ ). Recidivism was measured as the statistical probability of being convicted of a new alcohol-related traffic offense, including DWI, at two years post-entry. Proportional hazards survival analysis was performed on recidivism data extracted from State Department of Motor Vehicle (DMV) records.

Among the intent-to-treat sample (i.e., all individuals who initially entered the study), 5.4% of the DWI Court participants and 7.4% of the standard probationers were convicted of a new alcohol-related traffic offense. Although this difference was not statistically significant ( $p = .15$ ), it did reveal a marginal trend in the predicted direction. Among completers of their respective programs, the re-conviction rate was 3.6% for DWI Court graduates ( $n = 270$ ) and 6.4% for probation completers ( $n = 284$ ), which was statistically significant after controlling for the number of prior alcohol-related traffic offenses ( $p < .05$ ).

On one hand, these results are in support of the DWI Court. Among graduates, outcomes were significantly better than for probation completers. However, the magnitude of

this effect according to Cohen's (1988) criteria was small ( $h = .13$ ) and did not hold up for the original intent-to-treat sample. At a minimum, this requires replication in order for one to place confidence in the results.

It is also noteworthy that the evaluation was performed over a 63-month period beginning at or near the founding of the program. It is unclear what proportion of the sample entered the program after the DWI Court had been in operation long enough to develop and improve its services. If a sizeable proportion of the sample entered the program during its infancy year, this could have diminished the results to some degree.

#### *Rio Hondo DWI Court.*

Another randomized experimental evaluation was conducted of the Rio Hondo DWI Court in Los Angeles County, CA (MacDonald et al., 2007). Offenders convicted of a second or third misdemeanor DWI between May of 2000 and December of 2002 were eligible to participate. Consenting individuals were randomly assigned either to the DWI Court ( $n = 139$ ) or to standard adjudication ( $n = 145$ ). Outcomes included re-arrest rates for DWI and other alcohol-related crimes, as well as self-reported drunk-driving events and alcohol abuse. At the two-year follow-up, results revealed no significant differences on any outcome measure between the two conditions.

Importantly, the Rio Hondo DWI Court was created as an "experimental" program concurrently with the initiation of the research study (MacDonald et al., 2007, p. 9). It had not previously been in existence and ceased its operations upon conclusion of the research. There was apparently no opportunity for the program to pilot-test or improve its operations, nor was the outcome evaluation preceded by a process evaluation that documented the program's fidelity to applicable professional standards (NADCP, 1997; NDCI,

2006). Moreover, no dosage information was reported on how often the participants actually attended treatment sessions or status hearings or had been tested for substance use. As such, it is difficult to know whether the operations of this program were reflective of a typical DWI Court program.

It is also important to note that participants in the control condition received interventions that are ordinarily associated with a DWI Court, and not with probation as-usual—a confound known as “bleeding” or “contamination”. For example, the control participants were ordered to attend status hearings in court twice during the first six months of the program whereas DWI Court participants were ordered to attend status hearings only three times during the first six months (MacDonald et al., 2007, p. 11). The control participants were ordered to attend a total of five to seven court hearings whereas the DWI Court participants were ordered to attend a total of eight to ten hearings. This might have represented a negligible difference between the two groups on the one ingredient (court hearings) that most clearly distinguishes DWI Courts from other interventions for DWI offenders (e.g., Marlowe, 2006; Marlowe et al., 2004). It should not be surprising that outcomes were similar between the two groups because the probation subjects received key elements of the DWI Court model.

### *Multnomah DUI Intensive Supervision Program.*

Multnomah County, Oregon developed a court-supervised intensive probation program for felony and misdemeanor DWI offenders. Referred to as DISP (DUI Intensive Supervision Program), this program is primarily managed by the probation department but includes continuing court jurisdiction and court appearances at roughly four to six-month intervals. It is scheduled to be three years in length and incorporates a wide range of interventions, including intensive treatment and probation contacts, victim impact panels, electronic monitoring,

telephonic breath testing, driver's license suspension, mandatory vehicle sales and polygraph testing.

Outcomes from the DISP program ( $n = 460$ ) were compared to those of standard adjudication for DWI offenders drawn from neighboring counties ( $n = 497$ ) and matched on relevant baseline variables, including current age and number of prior DWI offenses (Lapham et al., 2006a?). Recidivism data were extracted from the state DMV and included the proportion of subjects convicted of a new DWI offense, driving with a suspended or revoked license, or moving traffic violation. The samples were drawn on a rolling basis from the start of the program in January, 1998 through March, 2001 and outcomes were evaluated through March, 2004. As such, outcomes were assessed between three and six years post-entry depending upon when a particular participant first entered the program.

Results revealed that DISP participants were less likely to be convicted of a subsequent DWI offense (9.8% vs. 18.3%), driving with a suspended or revoked license (14.6% vs. 27.2%) or traffic violation (28.0% vs. 38.4%). These effects were all statistically significant ( $p < .01$ ) and were in the small range ( $h = .22$  to  $.31$ ) according to Cohen's (1988) criteria. The time-delay until the first recidivist event was also significantly longer for DISP participants ( $p < .001$ ), suggesting they refrained from DWI conduct for a longer time after leaving the program.

While potentially supportive of the DWI Court model, it is difficult to know whether these superior effects were attributable to the court-based elements of the program as opposed to the exhaustive regimen of probationary interventions that were also available. The DISP program differed from standard probation on so many dimensions that it is not possible to determine which aspects may have elicited the beneficial effects. Indeed, against the backdrop of such an intensive and multifaceted program, it is often

difficult to experimentally isolate the effects of any one component. This could explain, in part, why the investigators were unable to detect specific effects for certain sanctioning elements of DISP, such as mandatory vehicle sales (Lapham et al., 2007).

### **Marginally Acceptable Evaluations of Mature Programs**

Two evaluations received marginally acceptable MQS scores and involved DWI Courts that had been in existence for an extended period of time prior to initiating the research.

#### *Georgia's Athens-Clarke, Chatham and Hall County DWI Courts.*

A multi-site evaluation was completed of three DWI Courts in Georgia (Meredith, 2007). Recidivism data were extracted from state criminal justice databases on re-arrests for any felony offenses, any misdemeanor offenses, drug offenses and DWI offenses at 12 and 24 months post-completion. Comparison samples were drawn from the same counties ( $n = 281$ ) prior to the founding of the DWI Courts. Although 645 offenders initially entered the DWI Court programs, analyses were only reported on 364 (56%) graduates.<sup>2</sup> Results revealed the DWI Court graduates had significantly fewer re-arrests in all offense categories at 12 months post-completion and in all offense categories other than drug crimes at 24 months post-completion.

Unfortunately, the failure to report outcomes on the entire intent-to-treat sample (i.e., on all individuals who initially entered the DWI Courts) renders the comparisons of

---

<sup>2</sup>A re-analysis of the data was completed in July of 2008, which included the entire intent-to-treat sample (Fell et al., 2008). The preliminary results, which have not yet been published, suggest there were superior outcomes for the DWI Court participants when the drop outs and terminated cases were included in the analyses.

questionable utility. It is not appropriate to select out successful cases from the DWI Court group (i.e., successful graduates) and compare them to the entire cohort of control subjects (e.g., Heck, 2006). Analyses should have been conducted on the intent-to-treat sample, or at a minimum the comparison group should have included only successful probation completers. Otherwise, there is a serious risk of a biased comparison unfairly favoring the DWI Court programs.

*Las Cruces, New Mexico DWI Court.*

A randomized experimental study was conducted in the Las Cruces, New Mexico DWI Court (Breckenridge et al., 2000). First and second-time DWI offenders who were determined by clinical staff to be alcohol dependent were randomly assigned either to the DWI Court ( $n = 39$ ) or to adjudication as usual ( $n = 36$ ). Recidivism data were obtained from the county Municipal Court database and included new convictions for traffic offenses as well as for alcohol, drug and other serious offenses at 24 months post-arrest.

Results revealed no significant differences between the two randomized groups in terms of convictions for any of the enumerated offenses. However, due to the small sample sizes in this study, there might have been insufficient statistical power to detect differences if they were present. In fact, approximately 15% of the DWI Court participants were convicted of an alcohol, drug or other serious offense, as compared to 22% of the control participants. This difference could reflect a small to moderate effect, but the sample sizes ( $n$ 's = 39 and 36) were only sufficient to detect relatively large effects (Cohen, 1988). With a larger sample size, this difference might have turned out to be statistically significant. At a minimum, it points to a potential trend favoring the DWI Court over adjudication as usual.

## DISCUSSION

This systematic literature review examined published and unpublished DWI Court program evaluations released through April 30, 2007. [See Addendum for a subsequent evaluation released in October 2007]. Although the results hint at emerging evidence potentially favoring the effects of DWI Courts, conclusions are seriously hampered by the disappointing state of the research in this area. A mere five evaluations were determined by independent raters to have employed good to marginally acceptable research methodology, and several of those evaluations still had serious flaws. These include: evaluating potentially immature programs, failing to conduct intent-to-treat analyses, and bleeding of the interventions across conditions.

In many ways, the current state of DWI Court research mirrors that of Drug Courts during the late 1990s to early 2000s when the U.S. Government Accountability Office (GAO, 2002) concluded that data were largely lacking to support the programs. Of course, an absence of data does not imply that a program is ineffective, as evidenced by a subsequent GAO report (GAO, 2005), several recent meta-analyses (Lowenkamp et al., 2005; Shaffer, 2006; Wilson et al., 2006) and numerous review articles (e.g., Aos et al., 2006; Belenko, 1998; Cissner & Rempel, 2005; Marlowe et al., 2003). All concluded that Drug Courts significantly reduce crime and drug use while participants are enrolled in the programs, and significantly reduce criminal recidivism following discharge. Unfortunately, until those data were appropriately collected and disseminated, Drug Courts remained vulnerable to serious criticisms, encroachments from competing philosophies, and funding cuts. At this juncture, DWI Courts are vulnerable to the very same criticisms and encroachments.

There is no alternative but to sponsor scientifically defensible studies that can fairly establish the effects of DWI



Courts as compared to adjudication as usual and as compared to alternative intervention approaches (e.g., intensive DWI probation). Because the methodological criteria outlined in Table 2 reflect best practices for the field of evaluation research, criteria such as these should be used to guide future designs of DWI Court program evaluations.

Most of the evaluations reviewed in this project provided insufficient information for determining how DWI Courts work and for what types of offenders. A large proportion (79%) of the evaluations failed to report any information on the dosages of services that were received by participants, such as the number of counseling sessions or status hearings that were attended. As a result, it was not possible to examine which services, if any, might have contributed to effective outcomes or how well the programs were implemented in practice. Future DWI Court evaluations should report on the dosages of services received by participants and conduct mediational analyses to determine which components contributed to effective outcomes.

The evaluations also generally limited outcome analyses to recidivism rates and graduation rates. Therefore, there was no way to examine effects on proximal or short-term outcomes, such as counseling attendance or abstinence rates, and to determine whether these proximal effects mediated longer-term outcomes. It is important to know, for example, whether reductions in alcohol or drug use lead to longer-term reductions in DWI recidivism, or whether these outcomes are relatively independent of one another. Future evaluations should report information on proximal outcomes and examine whether these proximal outcomes influence recidivism rates.

Virtually all of the evaluations reported recidivism rates over a specified time period following entry into or discharge from the programs. It would be useful to further break down recidivism events as having occurred either

during participants' active enrollment in the program or following graduation or termination. This would provide important information about whether outcomes tend to degrade after the period of intensive court supervision has ended.

Finally, it is important to recognize that no program would be expected to be effective for all DWI offenders. Drug Courts, for example, have been shown to be most effective for high-risk drug offenders characterized by more serious criminal backgrounds or treatment-refractory courses (Fielding et al., 2002; Lowenkamp et al., 2005; Shaffer, 2006). Alternative probation programs or treatment programs may be equally effective or more cost-efficient than Drug Courts for low-risk offenders (e.g., DeMatteo et al., 2006). DWI Courts might turn out to be necessary only for certain types of DWI offenders as well. Approximately one-half (57%) of the evaluations examined in this review reported client-level characteristics in their samples that are known to predict DWI recidivism and none conducted statistical analyses aimed at detecting potential interaction effects or moderator effects. If, in fact, DWI Courts are more effective for some types of DWI offenders but not others, failing to examine interaction effects could wash-out the results and lead to the unwarranted conclusion that DWI Courts are ineffective for the DWI population as a whole (e.g., Taxman & Marlowe, 2006).

A substantial literature base is available that identifies reliable and robust predictors of DWI recidivism. The most commonly identified DWI risk factors include current age, marital status, educational attainment, employment status, arrest BAC, number of prior DWI arrests, number of prior criminal arrests, alcohol use severity, and comorbid psychiatric disorders (e.g., Beerman et al., 1988; C'de Baca et al., 2001; Lapham et al., 2006b; Nochajski et al., 1993; Nochajski & Stasiewicz, 2006; Peck et al., 1993; Schell et al., 2006; Timken, 2002). These risk variables should be

carefully measured and examined in interaction analyses in future DWI Court program evaluations.

In addition, when non-randomized comparison samples are being used, it is incumbent upon the researcher to match the groups on at least some of these predictor variables, rather than simply matching on convenient demographic characteristics (e.g., race or county of residence) that may be easy to measure but do not necessarily relate to a risk for DWI recidivism. If matching is not feasible, then at a minimum it is necessary to statistically control for baseline differences between the study conditions on significant risk variables. Failing to do so renders the findings suspect and opens the study to the legitimate criticism that the “deck was stacked” from the outset in favor of the DWI Court program.

### **Limitations**

The primary limitation of this review relates to the coding procedures that were employed. The *Mesa Grande Coding System* was selected because it is commonly used in evaluations of substance abuse treatment interventions. Therefore, it provides a basis for comparing DWI Court evaluations against those of other substance abuse treatment programs. However, it could be argued that this coding system may be too strict or too lenient in terms of assessing the scientific integrity of program evaluations.

A relatively liberal cut-off score of 65% was set for “marginally acceptable” evaluations because most DWI Court evaluations are conducted in real-world programs by local evaluators and not in scientifically controlled research settings. Setting more stringent criteria could have the effect of excluding many evaluations that provide useful and practical information about how these programs perform in day-to-day practice. On the other hand, it is possible for evaluations having a single major flaw, such as failing to

include dropouts or terminated cases in the data analyses, to receive marginally acceptable scores using this liberal cut-off. Ultimately, each evaluation must be further assessed regarding its specific methodology, and some evaluations that received marginal scores may still need to be excluded from consideration on other grounds. Setting a relatively liberal cut-off score merely avoids prematurely excluding evaluations from further consideration.

As noted, the *Mesa Grande* system does require scientifically defensible evaluation designs that permit inferences of causality to be reached about the effects of the programs. Some DWI Courts may lack resources or scientific expertise to pass muster under this system. However, this is not to imply that their evaluations are useless. Although some evaluations may not be rigorous enough from a scientific perspective, they may still be acceptable and useful for local purposes, such as reporting process findings and cost data to funders or state or local governments. Being excluded from this systematic review should not be taken as an indication that evaluation findings have no value, and certainly should not be taken as an indication that the program itself is not effective. There are undoubtedly many effective programs that simply have not, as yet, been adequately studied.

Finally, there is room for debate about how to interpret the Program Maturity Index (PMI). As discussed earlier, it is generally viewed as preferable to give programs ample time to pilot-test their operations and implement indicated modifications before outcome analyses are conducted (Heck, 2006; Rempel, 2007). However, newer programs may also have certain advantages, such as motivated leadership, fresh political will, and new funding sources. It is possible that the effects of programs may degrade over the years as a result of reduced funding, changing political priorities, staff turnover or staleness of the operations. As such, it is not necessarily the case that

evaluations with low PMI's should be excluded from consideration or their results afforded less weight. Ideally, programs should be repeatedly evaluated over multiple years to permit a determination of whether outcomes tend to improve with experience or degrade from loss of interest or newer priorities.

### **Summary**

[2] In summary, although the results of this systematic review hint at emerging evidence potentially favoring the effects of DWI Courts, it is not possible at this juncture (as of 4/30/07) to reach scientifically defensible conclusions about the effects of DWI Courts due to the current state of the evaluation literature. It is hoped that the methodological criteria outlined in this review can serve as a template for future DWI Court program evaluations and assist practitioners and policymakers to become competent and effective consumers of DWI Court program evaluation findings.

**REFERENCES**

- Aos, S., Miller, M., & Drake, E. (2006). *Evidence-based public policy options to reduce future prison construction, criminal justice costs, and crime rates*. Olympia: Washington State Institute for Public Policy.
- Ball, S. A., Carroll, K. M., Canning-Ball, M., & Rounsaville, B. J. (2006). Reasons for dropout from drug abuse treatment: Symptoms, personality, and motivation. *Addictive Behaviors, 31*, 320–330.
- Becker, S. J., & Curry, J. F. (2008). Outpatient interventions for adolescent substance abuse: A quality of evidence review. *Journal of Consulting & Clinical Psychology, 76*, 531–543.
- Beerman, K. A., Smith, M. M., & Hal, R. L. (1988). Predictors of recidivism in DUIs. *Journal of Studies on Alcohol, 49*, 443–449.
- Belenko, S. (1998). Research on drug courts: A critical review. *National Drug Court Institute Review, 1*, 1–42.
- \*Breckenridge, J. F., Winfree, L. T., Maupin, J. R., & Clason, D. L. (2000). Drunk drivers, DWI “drug court” treatment, and recidivism: Who fails? *Justice Research & Policy, 2*, 87–105.
- C’de Baca, J., Miller, W. R., & Lapham, S. (2001). A multiple risk factor approach to predicting DWI recidivism. *Journal of Substance Abuse Treatment, 21*, 207–215.

- Cissner, A. B., & Rempel, M. (2005). *The state of drug court research: Moving beyond "do they work?"* New York, NY: Center for Court Innovation.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Compton, R., & Anderson, T. (1985, December). *The incidence of driving under the influence of drugs 1985: An update of the state of the knowledge* [DOT HS 806–900]. Washington, DC: National Highway Traffic Safety Administration.
- Cornish, J., & Marlowe, D. B. (2003). Alcohol treatment in the criminal justice system. In B. Johnson, P. Ruiz, & M. Galanter (Eds.), *Handbook of clinical alcoholism treatment* (pp. 197–207). Baltimore, MD: Lippincott, Williams & Wilkins.
- DeMatteo, D. S., Marlowe, D. B., & Festinger, D. S. (2006). Secondary prevention services for clients who are low risk in drug court: A conceptual model. *Crime & Delinquency*, 52, 114–134.
- DeYoung, D. J. (1997). An evaluation of the effectiveness of alcohol treatment, driver license actions and jail terms in reducing drunk driving recidivism in California. *Addiction*, 92, 989–997.
- Fell, J. C., Tippetts, A. S., & Langston, E. A. (2008, July). *An evaluation of the process and impact of three Georgia DUI Courts* [draft report]. Calverton, MD: Pacific Institute for Research and Evaluation.

- Festinger, D. S., Lamb, R. J., Marlowe, D. B., & Kirby, K. C. (2002). From telephone to office: Intake attendance as a function of appointment delay. *Addictive Behaviors*, 27, 131–137.
- Fielding, J. E., Tye, G., Ogawa, P. L., Imam, I. J., & Long, A. M. (2002). Los Angeles County drug court programs: Initial results. *Journal of Substance Abuse Treatment*, 23, 217–224.
- Freeman-Wilson, K., & Huddleston, C. W. (1999). *DWI/Drug Courts: Defining a national strategy*. Alexandria, VA: National Drug Court Institute.
- Greenfield, L. A. (1998). *Alcohol and crime: An analysis of national data on the prevalence of alcohol involved crime*. Washington, DC: U.S. Department of Justice, Office of Justice Programs.
- Heck, C. (2006, June). *Local drug court research: Navigating performance measures and process evaluations*. Alexandria, VA: National Drug Court Institute.
- Huddleston, C. W., Marlowe, D. B., & Casebolt, R. (2008, May). *Painting the current picture: A national report card on drug courts and other problem solving court programs in the United States* [Vol. II, No. 1]. Alexandria, VA: National Drug Court Institute.
- \*Jones, R. K. (2005, March). *Final report: Evaluation of the DUI Court program in Maricopa County, Arizona*. Winchester, MA: Mid-America Research Institute, Inc. of New England.



- Lapham, S. C., C'de Baca, J., Lapidus, J., & McMillan, G. P. (2007). Randomized sanctions to reduce re-offense among repeat impaired-driving offenders. *Addiction, 102*, 1616–1625.
- Lapham, S. C., C'de Baca, J., & McMillan G. P. (2006b). Psychiatric disorders in a sample of repeat impaired-driving offenders. *Journal of Studies on Alcohol, 67*, 707–713.
- \*Lapham, S. C., Kapitula, L. R., Baca, J. C., & McMillan, G. P. (2006a). Impaired-driving recidivism among repeat offenders following an intensive court-based intervention. *Accident Analysis & Prevention, 38*, 162–169.
- Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2005). Are drug courts effective?: A meta-analytic review. *Journal of Community Corrections, Fall*, 5–28.
- \*MacDonald, J. M., Morral, A. R., Raymond, B., & Eibner, C. (2007). The efficacy of the Rio Hondo DUI Court: A 2-year field experiment. *Evaluation Review, 31*, 4–23.
- Marlowe, D. B. (2006). Judicial supervision of drug-abusing offenders. *Journal of Psychoactive Drugs, SARC Suppl. 3*, 323–331.
- Marlowe, D. B., DeMatteo, D. S., & Festinger, D. S. (2003). A sober assessment of drug courts. *Federal Sentencing Reporter, 16*, 153–157.
- Marlowe, D. B., Festinger, D. S., & Lee, P. A. (2004). The judge is a key component of drug court. *Drug Court Review, 4*, 1–34.

- Marzuk, M. J., Tardiff, K., Leon, A. C., Stajic, M., et al. (1990). Prevalence of recent cocaine use among repeated heremotor vehicle fatalities in New York. *JAMA*, 263, 250–256.
- McCartt, A. T., Geary, L. L., & Nissen, W. J. (2002). *Observational study of the extent of driving while suspended for alcohol-impaired driving* [DOT HS 809 491]. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.
- \*Meredith, T. (2007, April). *Georgia's DUI Court programs reduce recidivism*. Applied Research Services, Inc.
- \*Michigan State Court Administrative Office & NPC Research. (2007, October). *Michigan DUI Courts outcome evaluation: Final Report*. Portland, OR & Lansing, MI: Authors. Available at: [http://www.npcresearch.com/Files/MI\\_DUI\\_Outcome\\_Evaluation\\_FINAL.pdf](http://www.npcresearch.com/Files/MI_DUI_Outcome_Evaluation_FINAL.pdf).
- Miller, W., Brown, J., Simpson, T., Handmaker, N., Bien, T., Luckie, L., et al., (1995). What works? A methodological analysis of the alcohol treatment outcome literature. In R. Hester & W. Miller (Eds.), *Handbook of alcoholism treatment approaches: Effective alternatives* (pp. 12–44). Needham Heights, MA: Allyn & Bacon.
- Miller, W. R., & Wilbourne, P. L. (2002). Mesa Grande: A methodological analysis of clinical trials of treatments for alcohol use disorders. *Addiction*, 97, 265–277.
- Moher, D., Schultz, K. F., & Altman, D. (2002). The CONSORT Statement. Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*, 285, 1987–1991.

- Moja, L. P., Telaro, E., D'Amico, R., Moschetti, I., Coe, L., Liberati, A., & The Metaquality Study Group. (2005). Assessment of the methodological quality of primary studies by systematic reviews: Results of the metaquality cross sectional study. *British Medical Journal*, 330, 1–5.
- National Association of Drug Court Professionals. (1997). *Defining drug courts: The key components*. Washington, DC: Office of Justice Programs, U.S. Department of Justice.
- National Drug Court Institute. (2006). *The Ten Guiding Principles of DWI Courts*. Alexandria, VA: Author
- National Highway Traffic Safety Administration. (1998). *Traffic safety facts 1997*. Washington, DC: Author.
- National Institute on Drug Abuse. (2005, August). *Drugged driving* [NIDA InfoFacts]. Washington, DC: Author.
- Nochajski, T. H., Miller, B. A., Wieczorek, W. F., & Whitney, R. (1993). The effects of a drinker-driver treatment program: Does criminal history make a difference? *Criminal Justice & Behavior*, 20, 174–189.
- Nochajski, T. H., & Stasiewicz, P. R. (2006). Relapse to driving under the influence (DUI): A review. *Clinical Psychology Review*, 26, 179–195.
- Peck, R. C., Arstein-Kerslake, G. W., & Helander, C. J. (1993). Psychometric and biographical correlates of drunk-driving recidivism and treatment program compliance. *Journal of Studies on Alcohol*, 55, 667–678.
- Popkin, C. L., & Wells-Parker, E. (1994). A research agenda for the specific deterrence of DWI. *Journal of Traffic Medicine*, 22, 1–14.

- Rempel, M. (2007). Action research: Using information to improve your drug court. In G. Berman, M. Rempel & R. V. Wolf (Eds.), *Documenting results: Research on problem-solving justice* (pp. 101–122). New York, NY: Center for Court Innovation.
- Robertson, R., Vanlaar, W., & Simpson, H. (2007, January). *10 steps to a strategic review of the DWI system: A guidebook for policymakers*. Ottawa, Ontario, Canada: Traffic Injury Research Foundation.
- Schell, T. L., Chan, K. S., & Morral, A. R. (2006). Predicting DUI recidivism: Personality, attitudinal, and behavioral risk factors. *Drug & Alcohol Dependence*, 82, 33–40.
- Schults, R. A., Elder, R. W., Sleet, D. A., et al. (2001). Reviews of evidence regarding interventions to reduce alcohol-impaired driving. *American Journal of Preventive Medicine*, 21, 66–88.
- Shaffer, D. K. (2006). *Reconsidering drug court effectiveness: A meta-analytic review*. Las Vegas, NV: Dept. of Criminal Justice, University of Nevada.
- Simpson, D. D., Joe, G. W., & Brown, B. S. (1997). Treatment retention and follow-up outcomes in the Drug Abuse Treatment Outcome Study (DATOS). *Psychology of Addictive Behaviors*, 11, 294–307.
- Simpson, H., Singhal, D., Vanlaar, W., & Mayhew, D. (2006, November). *The road safety monitor: Drugs and driving*. Ottawa, Ontario, Canada: Traffic Injury Research Foundation.

- Stark, M. J. (1992). Dropping out of substance abuse treatment: A clinically oriented review. *Clinical Psychology Review, 12*, 93–116.
- Taxman, F. S., & Marlowe, D. B. (Eds.) (2006). Risk, needs, responsivity: In action or inaction? [Special Issue]. *Crime & Delinquency, 52*(1).
- Timken, D. S. (2002). What works: Effective DWI interventions. In H. E. Allen (Ed.), *What works—Risk reduction: Interventions for special needs offenders* (pp. 27–89). Lanham, MD: American Correctional Association.
- U.S. Government Accountability Office. (2002). *Drug courts: Better DOJ data collection and evaluation efforts needed to measure impact of drug court programs* [No. GAO-02-434]. Washington, DC: Author.
- U.S. Government Accountability Office. (2005). *Adult drug courts: Evidence indicates recidivism reductions and mixed results for other outcomes* [No. GAO-05-219]. Washington, DC: Author.
- Vaughn, M. G., & Howard, M. O. (2004). Adolescent substance abuse treatment: A synthesis of controlled evaluations. *Research on Social Work Practice, 14*, 325–335.
- Wagenaar, A. C., & Maldonado-Molina, M. (2007). Effects of drivers' license suspension policies on alcohol-related crash involvement: Long-term follow-up in forty-six states. *Alcoholism: Clinical & Experimental Research, 31*, 1399–1406.

- Wagenaar, A. C., Zobeck, T. S., Williams, G. D., & Hingson, R. (1995). Methods used in studies of drink-drive control efforts: A meta-analysis of the literature from 1960 to 1991. *Accident Analysis & Prevention*, 27, 307–316.
- Wells-Parker, E., Bangert-Drowns, R., McMillen, R., & Williams, M. (1995). Final results from a meta-analysis of remedial interventions with drink/drive offenders. *Addiction*, 90, 907–926.
- Wilson, D. B., Mitchell, O., & MacKenzie, D. L. (2006). A systematic review of drug court effects on recidivism. *Journal of Experimental Criminology*, 2, 459–487.
- Yu, J. (2000). Punishment and alcohol problems: Recidivism among drinking-driving offenders. *Journal of Criminal Justice*, 28, 261–270.

\*Citations marked with an asterisk (\*) reflect DWI Court program evaluations that received Methodological Quality Scores (MQS) in the “good” to “marginally acceptable” range.

## **ADDENDUM**

### **Research Update: Michigan DUI Courts' Outcome Evaluation**

[3] This addendum summarizes the results of a three-county evaluation of DWI Courts conducted in the State of Michigan and released after the official cut-off date for the systematic review (Michigan State Court Administrative Office & NPC Research, 2007).

#### **Methods**

With funding from the Michigan Office of Highway Safety Planning, data for the evaluation were compiled by the Michigan Supreme Court State Court Administrative Office (SCOA), and outcome analyses were performed independently by NPC Research. Three DWI Courts located in Ottawa and Bay Counties and the City of Clarkston were evaluated. The Ottawa DUI Court serves individuals charged with a second DWI offense, the Bay County DUI Court serves second and third-time repeat DWI offenders, and the Clarkston DUI Court serves first-time and repeat DWI offenders.

The comparison samples consisted of DWI offenders from the same counties who would have been eligible for the DWI Courts, but had been arrested in the year prior to the founding of the programs. The comparison individuals were subjected to adjudication as usual and were commonly sentenced to probation. Outcomes were evaluated at one and two years post-entry to the DWI Court or to probation. For Clarkston County, recidivism data on felony and serious misdemeanor offenses were extracted from the Michigan State Police Criminal History Records Database and driving records were obtained from the Michigan Secretary of State. For the other two counties, recidivism data were extracted from the Michigan Judicial Data Warehouse, which includes

data on criminal arrests and drug or alcohol-related traffic offenses.

Outcome analyses were performed on an intent-to-treat basis including both graduates and unsuccessful terminations. Primary statistical analyses compared the percentages of individuals arrested for any new offense and the average number of arrests at one and two year follow-ups as well as DWI arrests at two-year follow-up. Survival analyses also compared the average length of time to the first arrest during the two-year follow-up period.<sup>3</sup>

## **Results**

The evaluation received a Methodological Quality Score (MQS) of 19 out of 23, satisfying 83% of recommended criteria. As such, it would have been included in the previous systematic review had the study been released prior to the cut-off date of April 30, 2007.

Table 5 presents re-arrest rates by county for the DWI Court and comparison samples as reported in the original evaluation report. Participants in DWI Court were significantly less likely in two out of the three counties to be arrested for any new offense within two years of entry, and significantly less likely to be arrested for a new DWI offense in one of the counties. In most of the comparisons, the trends favored better outcomes for the DWI Court participants; however, small sample sizes

---

<sup>3</sup> Within-group analyses involving only the DWI Court participants were also conducted. These analyses examined rates of positive drug and alcohol tests over time, compared outcomes between graduates and non-graduates, and identified predictors of successful completion. Because these analyses did not involve a comparison sample, they are not summarized in this update, but are available from the original evaluation report.



appear to have contributed to insignificant results in some instances due to inadequate statistical power. The estimated effect sizes (ES) ranged from  $h = .18$  to  $.57$ , which is in the small to moderate range according to Cohen's (1988) criteria, and most were between approximately 0.30 and 0.50. In many instances, however, the sample sizes only provided sufficient statistical power to detect large effects.

Table 5. Recidivism Outcomes for DWI Court Clients and Comparison DWI Probationers in Three Michigan Counties

	DWI Court	Comparison	<i>p</i>	ES
<b>Ottawa County</b>				
<i>1 year post-entry</i>	<i>n</i> = 143	<i>n</i> = 66		
% arrested	4%	15%	< .01	0.39
<i>2 years post-entry</i>	<i>n</i> = 72	<i>n</i> = 66		
% arrested	8%	24%	< .01	0.45
% arrested for DWI	1%	14%	< .05	0.57
<i>Table 5</i>				
<i>continues . . .</i>				

Bay County			
1 year post-entry	n = 85	n = 49	
% arrested	15%	22%	n.s. 0.18
2 years post-entry	n = 54	n = 47	
% arrested	18%	31%	< .05 0.30
% arrested for DWI	1%	6%	n.s. 0.29
Clarkston			
1 year post-entry	n = 89	n = 146	
% arrested	5%	6%	n.s. 0.04
2 years post-entry	n = 37	n = 144	
% arrested	5%	14%	n.s. 0.32
% arrested for DWI	2%	10%	n.s. 0.36

Notes: n.s. = not statistically significant. ES = effect size expressed as *h* for proportional data.

Similar differences were found when comparing the average numbers of arrests; however, those data are not presented in the interests of brevity. Some of the comparisons were not statistically significant presumably because the data were skewed (i.e., there were many zero values); however, the trends were virtually the same, favoring the DWI Court participants. In addition, survival analyses revealed DWI Court participants remained arrest-free for significantly longer periods of time than did the comparison probationers in two out of the three counties.

## **Conclusion**

Results of this study lend promising support for the DWI Court model. Given the limited research on DWI Courts, more high-quality evaluations are needed to confirm the effects of DWI Court programs. These evaluations are also needed to enhance practitioners' understanding of how DWI Court programs may exert positive effects and for which target populations they may be best suited.