



Published in final edited form as:

Psychol Addict Behav. 2008 March ; 22(1): 117–121. doi:10.1037/0893-164X.22.1.117.

The Test–Retest Reliability of the Form 90-DWI: An Instrument for Assessing Intoxicated Driving

Jennifer E. Hettema, William R. Miller, J. Scott Tonigan, and Harold D. Delaney
University of New Mexico, Department of Psychology

Abstract

Although driving while intoxicated (DWI) is a pervasive problem, reliable measures of this behavior have been elusive. In the present study, the Form 90, a widely utilized alcohol and substance use instrument, was adapted for measurement of DWI and related behaviors. Levels of reliability for the adapted instrument, the Form 90-DWI, were tested among a university sample of 60 undergraduate students who had consumed alcohol during the past 90 days. The authors administered the instrument once during an intake interview and again, 7–30 days later, to determine levels of test–retest reliability. Overall, the Form 90-DWI demonstrated high levels of reliability for many general drinking and DWI behaviors. Levels of reliability were lower for riding with an intoxicated driver and for variables involving several behavioral conjunctions, such as seat belt use and the presence of passengers when driving with a blood alcohol concentration above .08. Overall, the Form 90-DWI shows promise as a reliable measure of DWI behavior in research on treatment outcome and prevention.

Each year, there are an estimated 160 million incidents of driving while intoxicated (DWI; Quinlan et al., 2005), which result in over 17,000 deaths annually (National Highway Traffic Safety Administration, 2003). Effective DWI interventions are clearly needed; however, research on such interventions has been sparse (Penningroth, 1982; Vegega & Klitzner, 1998). One significant obstacle to research on treatment outcome is the lack of reliable measures of DWI behavior.

Investigators interested in researching DWI face the “dark-figure problem,” common to crime research, in which there is a high ratio of unknown to known incidents (Ross, 1993). Arrest/recidivism rates are the most commonly used measure of DWI, but surveys indicate that the ratio of actual DWI incidents to arrests is quite high (Hayslip, Kapusinski, Darbes, & Zeh, 1976) and may even exceed one arrest per 2,000 DWI episodes (Ross, 1993). In addition, various demographic factors, including gender, ethnicity, area of residence, and socioeconomic status, have been found to differentially influence arrest (Caetano & McGrath, 2005; Lapham, Skipper, Chang, & Kennedy, 1998; Vingilis, Adlaf, & Chung, 1982). For example, although White men have the highest rates of DWI, mixed race and Native American men are more likely to be arrested (Caetano & McGrath, 2005). Consequently, measurement of DWI using arrest/recidivism rates may result in a confounded picture of true DWI behavior.

Correspondence concerning this article should be addressed to Jennifer E. Hettema, who is now at the Department of Psychiatry, University of California, San Francisco, San Francisco General Hospital, Suite 7M, 1001 Potrero Avenue, San Francisco, CA 94110 jennifer.hettema@ucsf.edu.

This article is based on Jennifer E. Hettema’s master’s thesis, which was conducted at the University of New Mexico. Howard Delaney is Area Head of the Quantitative Methodology Program.

Other measures of DWI also suffer from methodological flaws. For example, roadside and other types of surveys often have significant refusal rates. As Ross (1993) has argued, “a high proportion of impaired drivers in, say, a five percent minority of refusers could greatly affect the impression of the extent of impairment in the group” (p. 777); high refusal rates would thus leave it impossible for researchers to make simple and uniform assumptions about the meaning of such estimates. In addition, refusal to participate in surveys has been found to be differentially related to important demographic factors, including youth, ethnicity, having a telephone, and having an outstanding arrest warrant (Lapham, Baum, Skipper, & Chang, 2000).

Survey data are typically collected in a quantity–frequency or a last-incident-recall format. Quantity–frequency measures have been criticized for their unreliability, and last-incident-recall measures can provide an unrepresentative sample of individuals’ behavior. In addition, both methods are dependent on subjective judgments of intoxication. Even when DWI is explicitly defined by use of a specific blood alcohol concentration (BAC), most individuals are poor estimators of their own BAC (Martin, Rose, & Obremski, 1991).

Lastly, hospital data are sometimes used in DWI research. Although BAC is supposed to be routinely collected in most hospitals, the procedure is often conducted at variable times after intake and sometimes is overlooked completely (Treno, Copper, & Roeper, 1994). In addition, physicians and emergency personnel often possess a great deal of latitude in determining who is tested, which results in nonrandom selection. In sum, arrest/recidivism rates, surveys, and hospital data provide information regarding DWI, but all of these measures suffer from a variety of methodological problems (U.S. Department of Health and Human Services and National Institute on Alcohol Abuse and Alcoholism, 2000). It was the purpose of the present investigation to identify and test the reliability of an instrument that is not limited to individuals who are being arrested or are injured, that is not dependent on subjective judgment of intoxication, and that aids individuals in accurately recalling a large sample of their behavior. The Form 90 (Miller & Del Boca, 1994) was identified as an instrument that could potentially help researchers avoid such pitfalls if adapted for DWI research. This instrument also seemed appropriate, given its consistently demonstrated reliability across a range of substance use behaviors (Tonigan, Miller, & Brown, 1997; Westerberg, Tonigan, & Miller, 1998) and populations (Scheurich et al., 2005; Slesnick & Tonigan, 2004) and its widespread administration, as evidenced by its frequent use within research on substance abuse outcomes, including several recent randomized multisite clinical trials (Project MATCH; Project MATCH Research Group, 1997; COMBINE; Anton et al., 2006).

The Form 90 (Miller & Del Boca, 1994) is a structured interview that combines the advantages of time line follow-back and consumption grid methods. Time line follow-back methodology has been previously applied to the measurement of DWI behavior, with promising results (Usdan, Schumacher, McNamara, & Bellis, 2002). The Form 90 utilizes a calendar to help individuals reconstruct a continuous stream of drinking behavior for the most recent 90 days. Previously tested and utilized versions of the Form 90 have not included any questions regarding DWI or riding with an intoxicated driver (RWID) behavior. The adapted version of the Form 90, the Form 90-DWI, follows the same format as did the original, with the exception that individuals are asked to recall specific aspects of DWI and RWID behaviors.

Method

Participants

On the basis of power analysis and the methodology of previous Form-90 reliability investigations, 60 participants were included in the present study. Participants were students recruited from upper and lower level undergraduate psychology courses at the University of New Mexico. A study announcement was posted on a bulletin board for viewing by psychology students who chose to participate in research for credit in their psychology courses. Inclusion criteria included being at least 18 years of age and having consumed alcohol in the past 90 days.

Participants, on average, were in their early 20s, with a mean age of 21.96 years ($SD = 5.84$ years), and were female ($n = 38$; 67.9%). More than half of the subjects were White (55.4%), 31.7% were Hispanic, 5.4% were Native American, and 5.4% were classified as other (Asian, Middle Eastern, or Indian). In addition, 42.9% of the participants were classified as freshmen, and they had a mean self-reported grade point average (GPA) of 3.21 ($SD = .48$). All had a valid driver's license, and 94.7% had a personal vehicle.

Measures

To answer adjunct questions not addressed within the present investigation, we administered several instruments to participants as part of the study, including a self-administered quantity–frequency measure of alcohol consumption, DWI behavior, and RWID behavior. However, the focus of the present investigation was on the reliability of the Form 90-DWI, which was administered on two occasions, 7–30 days apart.

Form 90-DWI—This semistructured interview assessed drinking quantity and frequency information, as well as DWI and RWID behaviors. Participants used a 90-day calendar, on which they identified general and personal holidays and memorable events, to facilitate memory and to reconstruct drinking day by day. For each day of drinking reported, participants were asked the type and amount of alcohol they had consumed and the period of time over which they had consumed the alcohol. Using this information, plus gender and body weight, we calculated estimated peak BAC via computer software (Markham, Miller, & Arciniega, 1993) that had good validity in prediction of actual BAC (Hustad & Carey, 2005). In addition, for each drinking day, information was gathered on whether driving had occurred, miles driven, the number of passengers present, and whether the participant's seat belt had been worn. We used computer projection to estimate drivers' BAC at time of travel. Information was also gathered regarding RWID, including the presence or absence of this behavior, the BAC of the driver (as calculated by the above-described procedures), miles driven, and whether the participant's seat belt had been worn. Figure 1 provides a flowchart that describes questions that were asked for each of the 90 days included in the calendar. The instrument can be found in its entirety at: <http://www.jennyhetteema.com/Form90.doc>.

Procedure

Participants admitted into the study provided informed consent and met individually with the investigators, who administered the Form 90-DWI and several other paper-and-pencil self-report instruments not included in the current investigation. The Form 90-DWI follows a semistructured interview format and takes approximately 45 min to administer. The interviewer recorded participant responses on the calendar and later calculated scores for the key variables reported in this study. Individuals returned for their follow-up assessment 7 to 30 days following this intake assessment, during which time they were readministered the Form 90-DWI. The modal number of days between assessments was 7, with a slightly

higher mean of 8.17 ($SD = 4.37$). The follow-up version of the Form 90-DWI measured the same period of time as did the intake version.

Follow-up data were always collected by a different interviewer, who was masked to intake data. All interviewers received extensive training by Jennifer E. Hetteema, who adapted the instrument and who has had extensive Form 90 training and experience. She taped and periodically reviewed all assessment sessions to ensure compliance with interview protocol.

Statistical Analyses

As has been done in previous investigations regarding the reliability of the Form 90 (Tonigan et al., 1997; Westerberg et al., 1998), intraclass correlation coefficients (ICCs) with a one-way random-effects model were used. ICCs are preferred when small sample sizes are being studied, because they provide a more conservative estimate of reliability. Compared with other estimates of reliability, such as Pearson correlations, that account only for the relative rank orderings of paired scores, ICCs consider the absolute differences between pairs of scores; this characteristic allows the variance under study to be partitioned into several components and controls for the amount of test–retest similarity that may be a product of chance alone (Cicchetti & Sparrow, 1981). The ICC method was chosen because it provides estimates of reliability that partition out the residuals component of systematic bias that may have been present in the sample (Shrout & Fleiss, 1979). We conducted analyses to determine the appropriateness of using this statistical test and found one violation of assumptions, with the error of several variables failing tests of normal distribution: namely, as reported levels of several variables increased, so too did their respective errors in measurement. This violation could have the effect of decreasing reliability estimates (Tabachnick & Fidell, 1996); however, given the nature of the investigation, Type II error was seen as less threatening than was Type I error. The more conservative estimate of reliability was chosen.

Results

Baseline Characteristics

Table 1 lists the baseline drinking, DWI, and RWID characteristics of the sample. Participants reported drinking 1–2 times per week, and average drinking episodes were at binge drinking levels (more than five drinks on a single occasion). These rates of drinking are higher than those observed among national adult samples (Naimi et al., 2003) but are consistent with rates of drinking reported among other university samples (Wechsler, Lee, Kuo, & Lee, 2000). Riding in a motor vehicle after drinking was common among the sample, with respondents reporting a mean of 4.6 ($SD = 8.0$) incidents of DWI and 1 ($SD = 1.6$) incident of RWID during the 90-day assessment period. Although variance was high for many of the tested variables, measures of central tendency were not significantly skewed by outliers.

Reliability Analyses

A 91.66% follow-up rate was achieved for the sample, and the ICCs for the 55 participants who received assessments at intake and at follow-up were calculated for the Form 90-DWI. Table 2 reports the test–retest reliability of these measures for drinking, DWI, and RWID. It should be noted that a sample of the studied population did not report engaging in some of the behaviors, such as DWI and RWID, at one or both assessment points. In these cases, no data were available for seat belt use, presence of passengers, or miles driven, hence the lower n for these analyses.

As shown in Table 2, ICC reliability estimates were high for Form 90-DWI measures of drinking quantity, frequency, and associated BACs. In addition, general measures of DWI were high, with the exception of several measures of DWI presence of passengers and seat belt use. These lower reliability estimates may be a by-product of the small sample sizes on which they were tested, or they may represent true measurement error for variables that are intended to capture the conjunction of several behaviors (e.g., drinking and driving and BAC >.08 and seat belt use). Measures of RWID showed less promise, with generally low levels of reliability.

We conducted additional reliability analyses to determine whether particular subgroups of the sample may have been responsible for driving the high correlations between test and retest scores. Because drinking distributions were highly skewed among the population, with many individuals engaging in low-frequency or low-quantity drinking, it was feared that light drinkers who engaged in very little DWI or RWID behavior may have been responsible for the high correlations observed. Consequently, we conducted median splits (*Mdn* = 4.15 drinks per day, *Mdn* = 85.6 percent days abstinent), using Form 90-DWI intake levels of drinking quantity and frequency, and recalculated ICCs for the top half of each group. As can be seen in Table 2, levels of reliability did not decrease significantly. It can be concluded that low-quantity/low-frequency drinkers did not inflate the reliability of measured variables.

Discussion

Our data indicate that the Form 90-DWI shows promise for providing a reliable estimate of drinking behavior and several important DWI behaviors, including frequency of DWI and associated DWI BACs. However, secondary DWI measures, such as seat belt use and presence of passengers and RWID behaviors, show less promise. Levels of reliability were lower for variables that involved a conjunction of several measured behaviors, though they were still at largely acceptable levels. In addition, as the BAC level associated with behaviors increased, reliability decreased. It is possible that this phenomenon is the result of memory deficits that have been documented to accompany high BAC levels.

The current study is limited by its generalizability. The sample size is small and is composed solely of university drinkers. However, the encouraging findings suggest that the instrument should be tested on larger, clinical samples. Assessing levels of validity is an important next step. This task may be difficult, however, as there is currently no established “gold standard” for assessment of DWI behavior. Recent use of alcohol ignition interlock recorders provides one promising avenue for collecting validity data.

In sum, the Form 90-DWI appears to yield reliable indices of DWI behavior among the tested sample, a finding that provides impetus for further research with additional samples. The Form 90-DWI shows promise for providing a much-needed measure of DWI intervention outcome, as currently utilized indicators, such as arrests and injuries, occur at low frequency and are influenced by a variety of confounding factors. The Form 90-DWI does not rely on the respondent’s subjective judgment of intoxication but rather estimates BAC from reconstructed drinking data. Although prevention of future arrests and injuries is an important objective, the goal of DWI interventions is presumably to reduce the occurrence of DWI. The current investigation provides preliminary evidence that the Form 90-DWI may be a reliable tool for assessing DWI behavior itself rather than just its tragic consequences.

References

- Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, et al. Combined pharmacotherapies and behavioral interventions for alcohol dependence: The COMBINE Study: A randomized controlled trial. *Journal of the American Medical Association*. 2006; 295:2003–2017. [PubMed: 16670409]
- Caetano R, McGrath C. Driving under the influence (DUI) among U.S. ethnic groups. *Accident Analysis & Prevention*. 2005; 37:217–224. [PubMed: 15667807]
- Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*. 1981; 2:127–137.
- Hayslip B, Kapusinski D, Darbes A, Zeh R. Evaluation of driving while intoxicated programs: Some methodological considerations. *Journal of Studies on Alcohol*. 1976; 37:1742–1746. [PubMed: 1003984]
- Hustad JTP, Carey KB. Using calculations to estimate blood alcohol concentrations for naturally occurring drinking episodes: A validity study. *Journal of Studies on Alcohol*. 2005; 66:130–138. [PubMed: 15830913]
- Lapham SC, Baum, Skipper BJ, Chang I. Attrition in a follow-up study of impaired drivers: Who is lost? *Alcohol and Alcoholism*. 2000; 35:464–470. [PubMed: 11022021]
- Lapham SC, Skipper BJ, Chang L, Kennedy R. Factors related to miles driven between drinking and arrest locations among convicted drunk drivers. *Accident Analysis and Prevention*. 1998; 30:201–206. [PubMed: 9450123]
- Markham MR, Miller WR, Arciniega L. BACCuS 2.01: Computer software for quantifying alcohol consumption. *Behavior Research Methods, Instruments and Computers*. 1993; 25:420–421.
- Martin CS, Rose RJ, Obremski KM. Estimation of blood alcohol concentrations in young male drinkers. *Alcoholism: Clinical and Experimental Research*. 1991; 15:494–499.
- Miller WR, Del Boca FK. Measurement of drinking behavior using the Form 90 family of instruments. *Journal of Studies on Alcohol*. 1994; (Suppl 12):112–118.
- Naimi TS, Brewer RD, Mokdad A, Denny C, Serdula MK, Marks JS. Binge drinking among U.S. adults. *Journal of the American Medical Association*. 2003; 289:70–75. [PubMed: 12503979]
- National Highway Traffic Safety Administration. *Traffic safety facts, 2002*. Washington, DC: U.S. Department of Transportation; 2003.
- Penningroth P. What works? Some observations about the evaluation of DUI programs. *Abstracts & Reviews in Alcohol & Driving*. 1982; 36:7–10.
- Project MATCH Research Group. Matching alcoholism treatment to client heterogeneity: Project MATCH post-treatment drinking outcomes. *Journal of Studies on Alcohol*. 1997; 58:7–29. [PubMed: 8979210]
- Quinlan KP, Brewer RD, Siegel P, Sleet DA, Mokdad AH, Shults RA, Flowers N. Alcohol impaired driving among U.S. adults, 1993–2002. *American Journal of Preventive Medicine*. 2005; 28:346–350. [PubMed: 15831339]
- Ross LH. Prevalence of alcohol impaired driving: An international perspective. *Accident Analysis and Prevention*. 1993; 25:777–779. [PubMed: 8297445]
- Scheurich A, Müller MJ, Angheliescu I, Lörch B, Dreher M, Hautzinger M, Szegedi A. Reliability and validity of the Form 90 interview. *European Addiction Research*. 2005; 11:50–56. [PubMed: 15608472]
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*. 1979; 86:420–428. [PubMed: 18839484]
- Slesnick N, Tonigan JS. Assessment of alcohol and other drug use by runaway youths: A test retest study of the Form 90. *Alcoholism Treatment Quarterly*. 2004; 22:21–34. [PubMed: 18563208]
- Tabachnick, BG.; Fidell, LS. *Using multivariate statistics*. 3. New York: Harper Collins; 1996.
- Tonigan JS, Miller WR, Brown JM. The reliability of the Form 90: An instrument for assessing alcohol treatment outcome. *Journal of Studies on Alcohol*. 1997; 58:358–364. [PubMed: 9203116]
- Treno AJ, Copper K, Roeper P. Estimating alcohol involvement in trauma patients: Search for a surrogate. *Alcoholism: Clinical and Experimental Research*. 1994; 18:1306–1311.

- U.S. Department of Health and Human Services and National Institute on Alcohol Abuse and Alcoholism. Tenth Special Report to the U.S. Congress on Alcohol and Health. Washington, DC: U.S. Government Printing Office; 2000.
- Usdan S, Schumacher JE, McNamara C, Bellis J. Measuring impaired driving behaviors among college students: Development and reliability of the Impaired Driving Assessment. *American Journal of Health Studies*. 2002; 18:10–18.
- Vegega M, Klitzner M. What have we learned about anti-drinking–driving programs? *Evaluation and Program Planning*. 1998; 11:203–217.
- Vingilis E, Adlaf EM, Chung L. Comparison of age and sex characteristics of police-suspected impaired drivers and roadside-surveyed impaired drivers. *Accident Analysis and Prevention*. 1982; 14:425–430.
- Wechsler H, Lee JE, Kuo M, Lee H. College binge drinking in the 1990s: A continuing problem. Results of the Harvard School of Public Health 1999 College Alcohol Study. *Journal of American College Health*. 2000; 48:199–210. [PubMed: 10778020]
- Westerberg VS, Tonigan JS, Miller WR. Reliability of Form 90D: An instrument for quantifying drug use. *Substance Abuse*. 1998; 19:179–189. [PubMed: 12511815]

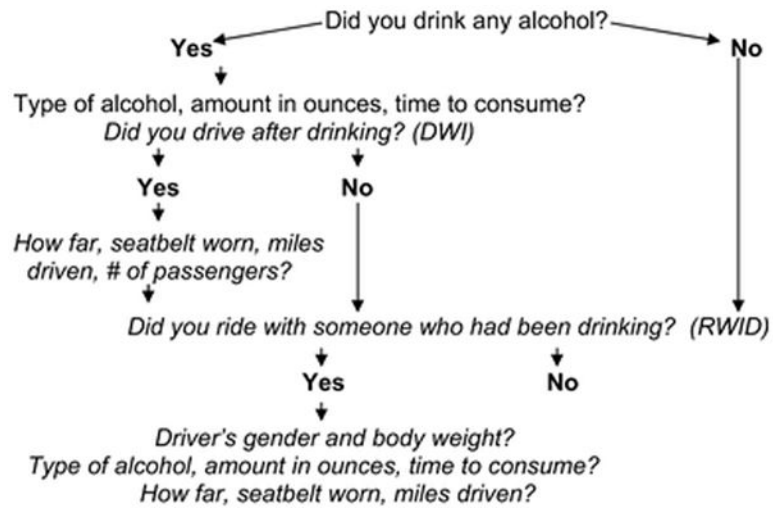


Figure 1. Items in italics represent additions to the original Form 90 scale. DWI = driving while intoxicated; RWID = riding with intoxicated driver.

Table 1

Intake Form 90-DWI Drinking, DWI, and RWID Levels

Variable	<i>n</i>	<i>M</i>	<i>SD</i>
% days abstinent	60	81.37	18.58
Drinks per drinking day	60	5.01	4.88
Highest BAC	60	0.22	0.21
Average BAC	60	0.09	0.08
No. DWI incidents	60	4.60	8.00
% drinking days with DWI incident	60	22.97	26.82
Peak DWI BAC	43	0.15	0.16
Average DWI BAC	43	0.09	0.10
% seat belt use during DWI incidents	43	94.65	20.37
% DWI incidents with passengers	43	54.83	41.12
Average miles driven for DWI incidents	43	9.59	9.28
No. DWI incidents with BAC >.08	60	2.10	5.41
% drinking days with DWI >.08 incidents	60	9.30	17.94
% seat belt use during DWI >.08 incidents	23	85.31	33.85
% DWI >.08 incidents with passengers	23	55.81	42.20
Average miles driven for DWI >.08 incidents	23	7.65	6.31
No. RWID incidents	60	1.00	1.60
% drinking days with RWID incident	60	7.59	11.79
% seat belt use during RWID incidents	25	89.74	29.47
Average miles driven for RWID incidents	25	12.46	14.27

Note. DWI = driving while intoxicated; RWID = riding with intoxicated driver; BAC = blood alcohol concentration.

Table 2

Form 90-DWI ICC Reliability Estimates for Entire Sample, Heavy Quantity Median-Split Sample (Mdn >4.15 DPD), and High-Frequency Median-Split Sample (Mdn >85.6 PDA)

Variable	Entire sample	Heavy quantity	High frequency
% days abstinent	.963* (n = 55)	.916* (n = 28)	.950* (n = 27)
Drinks per drinking day	.960* (n = 55)	.961* (n = 28)	.975* (n = 27)
Highest BAC	.907* (n = 55)	.886* (n = 28)	.924* (n = 27)
Average BAC	.897* (n = 55)	.865* (n = 28)	.962* (n = 27)
No. DWI incidents	.896* (n = 55)	.971* (n = 28)	.957* (n = 27)
% drinking days with DWI incident	.864* (n = 55)	.763* (n = 28)	.928* (n = 27)
Peak DWI BAC	.803* (n = 36)	.914* (n = 22)	.919* (n = 20)
Average DWI BAC	.765* (n = 36)	.787* (n = 22)	.823* (n = 20)
% seat belt use during DWI incidents	.943* (n = 36)	.952* (n = 22)	.951* (n = 20)
% DWI incidents with passengers	.598* (n = 36)	.515 (n = 22)	.660 (n = 20)
Average miles driven for DWI incidents	.737* (n = 36)	.912* (n = 22)	.856* (n = 20)
No. DWI incidents with BAC >.08	.898* (n = 55)	.935* (n = 28)	.942* (n = 27)
% drinking days with DWI >.08 incident	.924* (n = 55)	.868* (n = 28)	.937* (n = 27)
% seat belt use during DWI >.08 incidents	.590 (n = 10)	.712 (n = 16)	.594 (n = 14)
% DWI >.08 incidents with passengers	.551 (n = 19)	.450 (n = 16)	.682 (n = 14)
Average miles driven for DWI >.08 incidents	.799* (n = 19)	.830* (n = 16)	.917* (n = 14)
No. RWID incidents	.544* (n = 55)	.415 (n = 28)	.435 (n = 27)
% drinking days with RWID incident	.508* (n = 55)	.484 (n = 28)	.508 (n = 27)
% seat belt use during RWID incidents	.569 (n = 19)	.607 (n = 13)	.945* (n = 8)
Average miles driven for RWID incidents	.329 (n = 19)	.677 (n = 13)	.795 (n = 8)

Note. DWI = driving while intoxicated; ICC = intraclass correlation coefficients; DPD = drinks per day; PDA = percent days abstinent; BAC = blood alcohol concentration; RWID = riding with intoxicated driver.

* $p < .001$.