Development of Sobriety Tests for the Marine Environment

Dary D. Fiorentino, Bergetta M. Dietel, and Dulcemonica D. Jimenez

Six seated tests were evaluated in the laboratory to determine whether they would be feasible for use on the water as sobriety tests to measure impairment from alcohol at blood alcohol concentrations (BACs) of ≥0.08%. The standardized field sobriety tests (SFSTs) currently used at roadside are not suitable for the marine environment; marine law enforcement officers are left with insufficient methods to assess impairment on the water. One hundred fifty-seven participants were randomly assigned to a BAC group: 0.00%, 0.04%, 0.08%, and 0.12%. Six tests were administered to the participants by experienced law enforcement officers. Neither the testers nor the participants were privy to the participants' BACs. A variable called BAC status ($N = 138$) was obtained by dividing the average BAC into two groups: BAC < 0.08% and BAC ≥ 0.08%. A combination of four tests—horizontal gaze nystagmus (HGN), finger to nose (FTN), palm pat (PP), and hand coordination (HC)—correctly classified 82% of the BACs ≥0.08% and 67% of the BACs <0.08%, for an overall percentage correct of 72%. Four individual tests also predicted BAC status: HGN, FTN, PP, and HC. Four tests in combination and individually discriminated BAC status, although the overall percentages of accuracy, sensitivity, and specificity of the tests were below what was typically reported in literature on the roadside SFSTs. With the proper refinements, the four tests may assist marine officers with assessments of alcohol-related impairment in recreational boaters.

The roadside sobriety tests were developed in the late 1970s and early 1980s in two studies by the Southern California Research Institute (*1, 2*). The first study examined the usefulness of six candidate tests in detecting blood alcohol concentrations (BACs) of at least 0.10% (*1*). In that study, 238 participants were semirandomly assigned to one of four BAC Groups: 0.00%, 0.05%, 0.10%, and 0.15%. Law enforcement officers administered six tests to the participants. The six tests were one-leg stand (OLS), finger to nose (FTN), walk and turn (WAT), finger count (FC), tracing, and horizontal gaze nystagmus (HGN). On the basis of the results, the authors recommended a reduced battery of tests, which included the OLS, WAT, and HGN.

Nystagmus is a complex phenomenon that can occur for a variety of reasons (*3*). Within the impaired driving context, however, HGN specifically refers to a lateral jerking of the eyeball affected by alcohol, certain nervous system depressants, inhalants, and phencyclidine. The HGN test consists of six clues, three for each eye: lack of smooth pursuit, maximum deviation, and angle of onset (*4*). Four of six possible clues indicate impairment. The WAT test requires a person to assume a heel-to-toe position on a designated line, arms at the sides, and to listen while instructions are given. The person is then required to make nine heel-to-toe steps along the line, turn around keeping one foot on the line, and return with another nine heel-to-toe steps. Two of eight possible clues indicate impairment. The OLS test requires a person to stand, heels together, feet at a slight angle, and arms at the sides. The person is then required to raise one leg forward approximately 6 in. off the ground. Two of four possible clues indicate impairment.

In the second study (*2*), 297 participants were administered enough alcohol to reach peak BACs of 0.00%, 0.05%, 0.11%, and 0.15%. A combination of HGN, WAT, and OLS correctly identified 81.2% of the participants. Since the development of the roadside sobriety tests, they have been routinely used by law enforcement officers throughout the United States to identify BACs at or above the legal limit. Three validation studies have confirmed their usefulness (*5–7*).

The standardized field sobriety tests used at roadside to detect impairment in drivers with BACs of at least 0.08% are not suitable for the marine environment because two of the three tests (OLS and WAT) must be administered on a firm, flat surface. Marine officers who use these tests must bring the suspected boater to shore and wait a period of time, usually 15 min, to get the suspect adapted to being on land. Tests that can be administered without having to bring the suspect to shore will save time, but because of the motion of the boat on the water those tests would have to be administered with the suspect in a seated position. Previous efforts examined a variety of seated tests on boats and found encouraging results (*8, 9*).

The objective of this project was to develop sobriety tests that can be administered in the seated position to assist water patrol officers in detecting impairment caused by BACs of ≥0.08%. As in the roadside tests, the seated tests must be easy to administer, so as to not overburden law enforcement officers, who must continually monitor the environment for their own safety and the safety of the boaters suspected of impairment. The tests must discriminate impaired performance without the knowledge of the individual suspect's baseline performance. Most importantly, the tests must be useful for an arrest or release decision. Unlike the roadside tests, however, the seated tests cannot make use of any measure of equilibrium.

This paper reports the laboratory phase of the project. As was done in the past for the development of the roadside tests (*1, 4*), the usefulness of six candidate tests in detecting impairment was first evaluated in a controlled environment. Participants were tested at

Southern California Research Institute, 8115 Mammoth Avenue, Van Nuys, CA 91402. Current affiliations: B. M. Dietel, Seattle Institute for Biomedical and Clinical Research, 1660 South Columbian Way, S-116 ATC, Seattle, WA 98108; and D. D. Jimenez, Harder and Company Community Research, 315 West 9th Street, Suite 700, Los Angeles, CA 90015. Corresponding author: D. D. Fiorentino, dary.fiorentino@gmail.com.

Transportation Research Record: Journal of the Transportation Research Board, No. 2222, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 85–89.
DOI: 10.3141/2222-11

0.00% BAC, at half the legal limit (0.04%), at the legal limit (0.08%), and at 1.5 times the legal limit (0.12%).

The data were analyzed in five successive steps. First, researchers confirmed that the participants were tested at the intended target BACs. Second, researchers examined whether the total tests scores varied as a function of the four BAC groups. Third, researchers divided the BACs into groups and created a variable BAC status (BAC < 0.08% or BAC ≥ 0.08%) and examined whether the total tests scores varied as a function of it. Fourth, the correlations between BAC, BAC group, BAC status, and the six tests were examined. Finally, researchers conducted logistic regressions to establish whether the tests reliably predicted BAC status, individually and in combination.

METHOD

Participants

One hundred and fifty-seven men and women participated as paid volunteers. In addition, 17 scheduled participants failed to appear for testing, eight were dismissed for illegal drug use, and 13 were dismissed before testing because of evidence of health problems. The participants' ages ranged from 21 to 62 years (mean = 32.96, standard deviation = 10.70). Participants were 50.3% male and 49.7% female. They were 47.8% White, 20.4% African American, 3.2% Asian, 1.3% Pacific Islander, 0.6% American Indian, and 26.7% of other or unknown race. Twenty-eight percent of the participants were Latino. The participants' number of school years completed ranged from 1 to 24 (mean = 13.58, standard deviation = 3.63). Payment for participation was $100.

Testers

Twenty-four law enforcement officers participated in the study. Officers had an average of 9.7 years of experience administering the roadside SFSTs. Officers' participation spanned 4 days. Day 1 consisted of a training session on the tests' administration and scoring. Days 2, 3, and 4 were data collection days. Each data collection day lasted approximately 5 h. Officers were paid $100 per day.

Apparatus

The Intox EC/IR and the Alco Sensor FST (Intoximeters, Inc., St. Louis, Mo.) breath alcohol testing instruments were used to measure the participants' BAC.

Drug Screeners

All participants provided a urine specimen and were tested for drug use. Ten types of drugs were screened: methamphetamine, opiates, cocaine, marijuana, phencyclidene, benzodiazepines, barbiturates, methadone, tricyclic antidepressants, and amphetamine.

Pregnancy Tests

Female participants provided a urine specimen and the specimens were screened for human chorionic gonadotropin, the pregnancy hormone.

Tests

Six tests were evaluated:

- FTN. The FTN test required the participants to bring the tip of the index finger to touch the tip of the nose. It was performed with the eyes closed and the head tilted slightly back.
- Time estimation (TE). The TE test required the participants to estimate the passage of 30 s. It was performed with the eyes closed and the head tilted back. The test was scored as the absolute time deviation from 30 s.
- FC. The FC test required the participants to extend one hand forward palm up and to count to four while touching the tips of each finger with the tip of the thumb. The process was then reversed, and the participants counted backward. Three complete sets were performed.
- Hand coordination (HC). The HC test required the participants to perform a series of tasks with their hands. It was loosely adapted from the WAT test administered at roadside.
- Palm pat (PP). The PP test required the participants to extend one hand, palm up, and to place the other hand on it palm down. The participant was instructed to use the top hand to pat the bottom hand. The top hand rotated 180°, thereby alternating the pat between the back and the palm of the hand. The bottom hand remained stationary. The participant counted each pat aloud.
- HGN. Each eye was examined for lack of smooth pursuit, angle of onset, and jerking at maximum deviation.

Procedures

Participants were recruited with newspaper ads, Internet postings, flyers, and referrals. An initial telephone interview determined eligibility for the study. Applicants were screened in terms of health history, current health status, and use of alcohol and other drugs. The quantity-frequency–variability scale was used to classify applicants into five groups: abstainers, infrequent drinkers, light drinkers, moderate drinkers, and heavy drinkers (*10*). Only moderate and heavy drinkers were eligible to participate in the study. Pregnancy, chronic disease, or evidence of substance abuse resulted in exclusion from the study.

Participants were transported from their residence to the laboratory and from the laboratory to their residence by taxi or shuttle. Participants arrived at the facility in pairs at 9:00 a.m., 11:00 a.m., and noon. Thus, no more than six participants were tested per day.

On arrival at the laboratory, each participant gave informed consent to participate in the study, and each received a copy of the signed Informed Consent and of the Subjects' Bill of Rights. A breath alcohol test, a second administration of the Quantity-Frequency-Variability scale, a pregnancy test for women, and a drug screen confirmed eligibility for the study. Measurements of blood pressure, heart rate, height, and weight were taken next. Cardiovascular measures within acceptable ranges (systolic blood pressure = 120 ± 30 mmHg, diastolic blood pressure = 80 ± 20 mmHg, heart rate = 70 ± 20 beats/min) confirmed eligibility for the study.

Participants were randomly assigned to one of four groups (0.00% BAC, 0.04% BAC, 0.08% BAC, and 0.12% BAC) by lottery. No efforts were made to counterbalance moderate and heavy drinkers. Age, gender, weight, and height were used to calculate the alcohol dose. A drinking period of 30 min followed. Participants were served three equal-sized drinks at 10-min intervals and were instructed to pace each drink evenly over the entire 10 min. Research staff monitored the participants continually throughout the drinking period.

For participants in the 0.04% BAC, 0.08% BAC, and 0.12% BAC groups, the alcohol drink consisted of 1 part 80 proof vodka and 1.5 part orange juice. For participants in the 0.00% BAC Group, the placebo drink consisted of 1 part water and 1.5 part orange juice. The placebo glasses had their rim swabbed with vodka and 10 mL of vodka floated in each of them to produce an initial taste and odor of alcohol.

Twenty minutes after the end of the third drink, BAC measurements were obtained at 5-min intervals until the peak BAC was detected. Peak BAC was expected 30 min after the end of the third drink. For the 0.00% BAC group, testing occurred at the first available testing window 30 min after drink. Participants were not privy to their target BAC.

When the participants reached the target BAC on the descending limb of their BAC curve, they were brought in the testing room and were asked to sit down. The battery of tests was administered twice to each participant, each time by a different tester, with only one tester in the testing room at a time. Only the results from the first battery were compiled. The second battery was for practice only, because participants' learning affected the test scores.

Testers remained in a separate room and had no interaction with the participants before testing them. A staff member was present during testing to ensure that the interaction between participant and tester was limited to the administration and scoring of the tests. A BAC reading was obtained immediately after the testing.

BAC and test order were counterbalanced. When the participants' BAC dropped below 0.03%, they were debriefed, paid $100, and transported home by taxi.

RESULTS

BACs

Of the 157 participants, 39 were assigned to the 0% BAC condition, 40 to the 0.04% BAC condition, 39 to the 0.08% BAC condition, and 39 to the 0.12% BAC condition (see Table 1). Unequal group sizes were the result of some participants' failure to meet study criteria. Moderate and heavy drinkers were equally divided among the four BAC groups.

In general, the testing BACs were slightly lower than the target BACs, for two reasons. First, the dosing procedure was aimed at avoiding overdosing the participants, for obvious health and safety reasons. Second, a bottleneck occasionally resulted when two participants reached the target BAC at the same time, which delayed some of the testing. The following analyses were conducted with the average of the pretest BACs and the posttest BACs.

TABLE 1 BAC and Average Total Test Scores by BAC Group

Variable	BAC Group 1	2	3	4	*F*
BAC(%)					
Mean	0.000	0.038	0.079	0.110	1,527.37***
SD	0.000	0.009	0.004	0.012	
N	39	40	39	39	
FTN					
Mean	7.16	8.65	9.44	11.38	5.13**
SD	4.06	4.23	4.80	5.99	
N	38	40	39	39	
TE					
Mean	6.49	7.48	7.49	7.95	0.35
SD	5.30	8.61	5.14	6.29	
N	39	40	39	39	
FC					
Mean	4.74	6.90	7.46	7.90	2.19
SD	4.64	6.56	5.77	6.45	
N	39	40	39	39	
HC					
Mean	2.64	2.40	2.74	3.59	2.91*
SD	1.86	1.96	1.83	1.96	
N	39	40	39	39	
PP					
Mean	1.51	1.65	1.67	2.62	5.12**
SD	1.14	1.59	1.06	1.70	
N	39	40	39	39	
HGN					
Mean	2.03	3.60	4.85	5.28	23.53***
SD	2.02	2.23	1.74	1.43	
N	39	40	39	39	

NOTE: SD = standard deviation. BAC groups 1 to 4 had target BACs of 0.00%, 0.04%, 0.08%, and 0.12%, respectively, on the descending limb of the BAC curve. Lower scores indicate better performance.
$*p \le .05$, $**p \le .01$, $***p \le .001$.

Differences Across Four BAC Groups

The mean scores for the six tests across the four BAC groups increased with higher BACs. Mean scores increased with higher BACs. The mean score differences across BAC groups were statistically significant for FTN, HC, PP, and HGN (Table 1).

Differences Across BAC Status

In the field, marine officers need to assess whether boaters' BACs are ≥0.08%. Thus, it is important to examine whether the mean scores for the six tests differ significantly between two conditions, BAC < 0.08% or BAC ≥ 0.08%. To that end, a new variable was created by dividing the average BAC of the test battery into two categories: BACs < 0.08% and BACs ≥ 0.08%. Characteristics of the resulting variable (BAC status) are shown in Table 2. BAC status analyses were based on 138 of the 157 cases, because 19 participants had pretest and posttest BACs that were on both sides of the 0.08% cutoff. The mean scores for the six tests across BAC status are shown in Table 2. The mean scores across BAC status were consistent with the results across BAC Groups.

Correlations

TE and FC did not reliably correlate with BAC, BAC group, or BAC status. The correlations between HGN and BAC, BAC group, and BAC status were .56, .55, and .44, respectively (all with $p \le .01$). The correlations between FTN and BAC, BAC group, and BAC status were .29, .30, and .25, respectively (all with $p \le .01$). The correlations between PP and BAC, BAC group, and BAC status were .24, .26, and .26, respectively (all with $p \le .01$). The correlations between HC and BAC, BAC group, and BAC status were .19, .18, and .19, respectively (all with $p \le .05$).

TABLE 2 BAC and Average Total Tests Scores by BAC Status

	BAC Status		
Variable	BAC < 0.080%	BAC ≥ 0.080%	*F*
BAC(%)			
Mean	0.023	0.102	463.16***
SD	0.024	0.016	
N	85	53	
FTN			
Mean	8.11	10.77	9.37**
SD	4.29	5.87	
N	84	53	
TE			
Mean	6.91	7.91	0.75
SD	6.90	6.14	
N	85	53	
FC			
Mean	5.98	7.98	3.62
SD	5.73	6.45	
N	85	53	
HC			
Mean	2.54	3.28	4.89*
SD	1.89	1.96	
N	85	53	
PP			
Mean	1.64	2.42	9.49**
SD	1.36	1.57	
N	85	53	
HGN			
Mean	2.98	5.06	32.96***
SD	2.28	1.67	
N	85	53	

NOTE: Lower scores indicate better performance.
*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.

BAC Status Classifications

The question of how the tests would predict BAC status was addressed next. Only the four tests that were found to have statistically significant differences across BAC status were included in these analyses. Note that because the prediction analyses build on the previous analyses, they may capitalize on chance. The following results, therefore, must be interpreted with caution.

Combined Tests

Logistic regression was used to predict BAC status with HGN positive or negative, FTN total score, PP total score, and HC total score as the predictors. A test of the full model with the four tests against a constant-only model was statistically significant, χ^2 (4, N = 137) = 33.89, $p < .001$. As shown in Table 3, the combination of the four tests correctly classified 82% of the BACs ≥ 0.08%, 67% of the BACs < 0.08%, for an overall percentage correct of 72.3%. Of the individual tests, however, only HGN positive or negative reliably predicted BAC status, χ^2 (1, N = 137) = 16.13, $p < .001$, indicating that FTN, PP, and HC did not improve the prediction beyond that of HGN.

Horizontal Gaze Nystagmus

A test of the full model with HGN positive or negative scores (negative = three or fewer clues, positive = four or fewer clues) against a constant-only model was statistically significant, χ^2 (1, N = 138) = 26.48, $p < .001$. As shown in Table 3, HGN alone correctly predicted BAC status in 67.4% of the cases.

Finger to Nose

The positive or negative criterion for FTN was set at nine clues based on analyses from pilot data not reported here. With that criterion, a test of the full model with FTN against a constant-only model was statistically significant, χ^2 (1, N = 137) = 4.38, $p < .05$. FTN alone correctly predicted BAC status in 59.9% of the cases.

Palm Pat

The positive or negative criterion for PP was set at two clues based on analyses from pilot data not reported here. With that criterion, a test of the full model with PP against a constant-only model was statistically significant, χ^2 (1, N = 138) = 4.23, $p < .05$. PP correctly predicted BAC status in 57.2% of the cases.

Hand Coordination

The positive or negative criterion for HC was set at three clues based on analyses from pilot data not reported here. With that criterion, a test of the full model with HC against a constant-only model was statistically significant, χ^2 (1, N = 138) = 3.87, $p < .05$. HC correctly predicted BAC status in 57.2% of the cases.

DISCUSSION OF RESULTS

The objective of this project was to develop sobriety tests for the marine environment. Six seated tests were evaluated in the laboratory to determine their feasibility for use on the water. Data were obtained

TABLE 3 Summary of Results from Tests

Test	Prevalence	% Correct	Sensitivity	Specificity	LR+	LR−
Combination	0.39	72.3	81.1	66.7	2.43	.28
HGN	0.38	67.4	86.8	55.3	1.94	.24
FTN	0.39	59.9	58.5	60.7	1.48	.68
PP	0.38	57.2	66.0	51.8	1.37	.66
HC	0.38	57.2	64.2	52.9	1.36	.68

NOTE: LR+ = positive likelihood ratio; LR− = negative likelihood ratio.

under double-blind conditions, at relatively low BACs. The six tests were administered by law enforcement officers with an average 9.7 years of experience administering the roadside SFSTs.

The combination of four tests, HGN, FTN, PP, and HC, correctly predicted BAC status in 72.3% of the cases. The positive likelihood ratio of 2.43 and the negative likelihood ratio of 0.28 indicate that the combined tests are useful in detecting alcohol-related impairment, but not conclusively.

The overall correct percentages, sensitivity, and specificity of the tests were below what is typically reported in the literature on the roadside SFSTs. Comparisons with prior studies, however, should be made with caution. First, in this study, the average BACs were considerably lower than in previous studies. In the Burns and Moskowitz study, for example, 48 participants were tested at a mean BAC of 0.120%, and 16 participants were tested at a mean BAC of 0.156% (*1*).

In comparison, in the current study, the highest BAC group was tested at a mean BAC of 0.110%. The wider distribution of BACs in the previous studies may have made the impairment or no impairment decision less difficult than in the current study. Second, the impairment or no impairment decisions were made exclusively on the basis of the tests, without external clues such as smell of alcohol, appearance, speech, and demeanor. Third, although the officers in the current study were required to have prior experience administering the roadside SFSTs, and were, therefore, assumed to have nearly equal proficiency in administering HGN, great differences in proficiency were in fact observed between the officers. Five of the 24 study officers had overall percentage correct for HGN of less than 50%. In addition, given that officers collected data for 3 days, with six participants scheduled per day, the maximum number of participants that could be examined by a single officer was 18, which may not have been enough to master the tests. In retrospect, it appears that the issue of officer proficiency was not given proper consideration in this study. Future studies should set up proficiency criteria for officers' participation, improved training, and asymptotic test performance before data collection.

Although the tests, as administered and scored by officers in the laboratory, had lower correct percentages, sensitivity, and specificity than is typically reported in the literature, they showed enough promise to warrant a field study. The field study, reported elsewhere, was conducted on the water by highly trained marine officers accompanied by civilian observers (*11*). The results indicated that the overall correct percentages, sensitivity, specificity, and reliability of the tests on the water were consistent with what is typically reported in the literature on roadside sobriety tests. Thus, the four tests may assist well-trained marine officers with assessments of alcohol-related impairment in boaters.

ACKNOWLEDGMENTS

This research was supported by funding from the U.S. Coast Guard and by the National Association of State Boating Law Administrators.

REFERENCES

1. Burns, M., and H. Moskowitz. *Psychophysical Tests for DWI Arrest.* DOT HS-802-424. U.S. Department of Transportation, 1977.
2. Tharp, V., M. Burns, and H. Moskowitz. *Development and Field Test of Psychophysical Tests for DWI Arrest.* DOT-H-805-864. U.S. Department of Transportation, 1981.
3. Dell'Osso, L. F. Nystagmus and saccadic intrusions and oscillations. *Current Neuro-Ophthalmology,* Vol. 2. Year Book Medical Publishers, Chicago, Ill., 1990, pp. 147–182.
4. National Highway Traffic Safety Administration. *Horizontal Gaze Nystagmus: The Science and the Law.* DOT-HS-808-938. U.S. Department of Transportation, 1999.
5. Burns, M., and E. W. Anderson. *A Colorado Validation Study of the Standardized Field Sobriety Test (SFST) Battery.* Colorado Department of Transportation, Aspen, 1995. http://www.bobkeeferlaw.com/library/1995_Colorado_Validation_Study___Final_Report.pdf. Accessed April 13, 2011.
6. Burns, M., and T. Dioquino. *A Florida Validation Study of the Standardized Field Sobriety Test (SFST) Battery.* Florida Department of Transportation, Tallahassee, 1997. http://www.duianswer.com/library/1997_Florida_Validation_Study_of_SFST___Burns___Dioquino.pdf. Accessed April 13, 2011.
7. Stuster, J. Validation of the Standardized Field Sobriety Test Battery at 0.08% Blood Alcohol Concentration. *Human Factors,* Vol. 48, No. 3, 2006, pp. 608–614.
8. Sussman, E. D., A. Needalman, and P. H. Mengert. *An Experimental Evaluation of a Field Sobriety Test Battery in the Marine Environment.* U.S. Coast Guard, U.S. Department of Transportation, 1990.
9. McKnight, A. J., J. E. Lange, and A. S. McKnight. Development of a Standardized Boating Sobriety Test. *Accident Analysis and Prevention,* Vol. 31, No. 1, pp. 147–152.
10. Cahalan, D., I. H. Cisin, and H. M. Crossley. *American Drinking Practices: A National Study of Drinking Behavior and Attitudes.* Rutgers Center of Alcohol Studies, New Brunswick, N.J., 1969.
11. Fiorentino, D. Validation of Sobriety Tests for the Marine Environment. *Accident Analysis and Prevention,* Vol. 43, No. 3, pp. 870–877.

The Marine Safety and Human Factors Task Force peer-reviewed this paper.