



DWI

BY ALLEN M. TRAPP JR.

Field Sobriety Tests: The NHTSA Numbers They Don't Want You to Know

A Quick Trip Back in Time

For several decades, law enforcement officers have employed a variety of field sobriety tests to determine if a person is driving under the influence of alcohol. Once upon a time they were looking for drunk drivers, but as we all know, it has been over a quarter of a century since Jimmy Carter was president. Twenty or 30 years ago, it was common for Georgia State Troopers to have motorists blow into their "Smokey Bear" hats to guesstimate how much alcohol the motorists had consumed. Needless to say, some in the law enforcement community wanted something more.

Back in the 1970s (and in some jurisdictions to this very day) a wide variety of field sobriety tests existed, ranging from blowing in the hat to tracing on paper. Beginning in 1975, studies were sponsored by the National Highway Traffic Safety Administration (NHTSA) through a contract with the Southern California

Research Institute (SCRI) to determine which of the field sobriety tests were the most accurate. In other words, SCRI was not charged with developing tests but only with evaluating those already in use.

NHTSA, a branch of the Department of Transportation (DOT), requested proposals in the mid-1970s to conduct research on identifying the best "field sobriety tests" that an officer could use at roadside. SCRI Director Marcelline Burns, a research psychologist, and her group created a technical and cost proposal. NHTSA awarded SCRI a contract to do the research, and the first study reported in 1977.¹ The work by Burns on the 1977 study began in 1975 when she, after a literature search of related material, participated in ride-a-longs with police officers all over the United States. Burns developed a list of 16 tests thought to be feasible as potential "sobriety" tests. Using those tests, SCRI conducted some pilot tests with a small group of people and selected six tests for the 1977 study.

The 1977 SCRI report recommended the use of three of the six tests, namely, the walk-and-turn, the one-leg stand, and the horizontal gaze nystagmus (referred to as "alcohol gaze nystagmus" in the 1977 final report). The other tests used in the study were the finger-to-nose, the finger count, and the tracing test. The Romberg test, alphabet test, and subtraction tests were interchangeably used.

The testing, all of which was conducted in a laboratory setting, lasted approximately one year. It involved 238 drinking subjects and 10 police officers. The individuals selected as subjects for the research were licensed drivers and alcohol consumers. They were instructed not to eat for four hours before they were given measured doses of alcohol; however, the subjects did not know the amount of alcohol they consumed. Their breath alcohol concentrations (BAC) were measured, and then they were subjected to the six tests listed above.

The 1977 study concluded that the Romberg test and the finger-to-nose test merely reflected the presence of alcohol, but "did not increase the predictive ability of testing." In other words, the finger-

to-nose and Romberg tests did not add anything to the predictability of a subject's level of intoxication. It is also interesting to note that the finger count, finger-to-nose, Romberg, alphabet, and tracing tests were not recommended for use as sobriety tests. Although they were used and were part of the 1977 NHTSA study, none were selected as being indicators of anything, let alone as indicators of intoxication.

Some interesting statistics came out of the 1977 study. The error rate of the 10 officers involved in the study was of primary significance. *Their error rate was an astounding 47 percent!* People interested in the error rate will have to read the report because *this information cannot be found in an NHTSA manual.* That is to say, in the 1977 study the officers made the decision to "arrest" a total of 101 people. Of those people arrested, 47 percent had a BAC under 0.10 percent.² This was totally unacceptable, even according to the authors of the study. Marcelline Burns later tried to attribute the high error rate to the inexperience of the officers used in the study.

If this were true, it would seem inexplicable that the researchers would again use inexperienced officers in the 1981 NHTSA study. It is significant to note that approximately 80 percent of the subjects used in the 1977 study were in their 20s, and about two-thirds of them were male.³ Perhaps it is just the author's experience, but muscle tone and physical dexterity began deteriorating within a year or two following his 10th high school reunion.

The 1981 Study⁴

The 1977 study recommended further review, and NHTSA awarded SCRI a second contract for retesting and standardization. This second study resulted in the 1981 NHTSA report. In the 1981 study only the three-test battery was used. The 1981 study, like the 1977 study, was done only in a laboratory setting, except for a handful of experiments conducted at the end of the study. Burns states that the officers again made their decisions to arrest or not to arrest based on the pre-

diction that the subject's BAC was over or under 0.10 percent. There were 296 subjects in the 1981 study.

Some "divided attention" components were added midstream during the 1981 study. For example, Burns describes a divided attention component of the walk-and-turn test as the portion of the test wherein the subject is requested to stand with one foot in front of the other on the line while listening to the instructions. This is what officers nowadays refer to as the "instructional phase." The standardization effort was motivated by a desire to establish a consistent method of administering the tests, giving instructions, demonstrating the tests, and scoring. The goal was to ensure that if an officer in Tennessee administers the field sobriety tests and an officer in Montana also gives the tests, the two officers would arrive at the same conclusions. Burns considered the order in which the tests were given to be irrelevant.

In the 1981 study, *out of 118 decisions by the officers to arrest, 32 percent of them were wrong.*⁵ This is only slightly better than the 1977 study, which had a 47 percent error rate of false arrests. Also, in the 1981 study, *the officers misjudged as impaired 18 percent of the subjects who had no alcohol in their system.*⁶ Burns makes a rather twisted attempt at explaining this result. She opines that the study was done "next to the drug capital of the world." In other words, she hints that since none of these people were screened for drugs, they may have been impaired on substances other than alcohol. This simply is unsound logic. If one were to accept her logic, perhaps this could be grounds to invalidate the entire study since none of the subjects, including those who had ingested alcohol, had been screened for drugs. *The officers in the 1981 project believed 31 percent of the people who were at a 0.05 percent BAC to be impaired.*⁷

The most interesting statistics from the 1981 study, however, as discussed by Cole and Nowaczyk,⁸ involve the "dosing differential" of the subjects tested. Most of the subjects (78 percent) were dosed with either high BAC (about 0.15 percent) or low BAC (0.05 percent and below).⁹ These should have been easy decisions since, as a practical matter, it should have been easy for the officers to score an individual as being above a 0.10 percent BAC when they are 0.15 percent BAC and above. The same would be true of someone 0.05 percent and below. NHTSA claims an overall accuracy rate of 80 percent when using the three-test battery; however, this overall accuracy rate of

80 percent is questionable when over two-thirds (78 percent) should be considered "gimmies" (either dosed high or low, hence the "dosing differential"). In other words, the data of the individuals dosed between 0.05 percent and 0.15 percent would undoubtedly have a lower accuracy rate; however, that data is unavailable. Cole and Nowaczyk opine that one factor in determining the "improval" of the false arrest numbers (47 percent in 1977 down to 32 percent in 1981) could be due in part to the dosing differential.

The number of subjects dosed in the mid-range (0.05 percent to 0.15 percent) went down from 27 percent¹⁰ in the 1977 study to 22 percent in the 1981 study. What this means is that only 22 percent of the subjects in the 1981 study were in the more difficult to determine range between 0.05 percent and 0.15 percent BAC. Researchers involved in the 1981 study claim a "reliability study" as part of their research. Reliability basically refers to consistency, or the ability to get the same results each time. The reliability portion consisted of asking 145 of the subjects to come back for retesting two weeks after the original study. The "reliability factor" was 0.77. This "reliability correlation coefficient" is based on a scale from almost zero to 1.00. Note that a correlation coefficient of 0.9 or above is expected for academic reading tests such as the SAT. This inter-rater reliability coefficient dropped to 0.57¹¹ when done by different officers. Thus, when different officers tested the same subjects at the same dose level, the reliability level was pathetic, and far below scientific acceptability.

The age and gender of the subjects used in the 1981 project, as with the 1977 study, are highly significant when considering any interpretation of the results. In the 1981 study, a whopping 80 percent of the subjects were between the ages of 21 and 34. Again, as with the 1977 study, about two-thirds of the subjects were male.¹² The use of a predominately male population in their 20s means that observers should question the applicability of the test results to the population as a whole.

The Good-Augsberger Study

Two optometrists at Ohio State University, Gregory W. Good and Carol R. Augsberger, published one of the earliest non-NHTSA studies of field sobriety tests in the American Journal of Optometry & Physiological Optics.¹³ This 1986 article is highly complimentary of the standardized field sobriety test (SFST)

program at the Ohio State Highway Patrol Academy and regurgitates NHTSA statistics without any critical analysis. The article dutifully reports that 92 percent of subjects scoring four "points" or higher on the HGN registered BACs above .10 percent.

The authors, however, overlooked the bad news – false positives – although their own charts published with the article reveal a startling rate. Fully 81.5 percent of those with BACs under .10 percent also demonstrated four or more clues.¹⁴ Although NHTSA trumpeted that the exercise is "92 percent accurate in identifying intoxicated people," there was a concerted effort to ignore the fact that the data says the test is 82 percent inaccurate as applied to *innocent people*.

The Field Validation Studies

The three-test battery of SFSTs has been promoted by NHTSA over the past 20 years and has been adopted by all 50 states. In three highly publicized "validation studies," NHTSA claims to have found the proof that these FSTs are valid measures of BAC. All of the field studies are pretty consistent in terms of low false negative

David M. Benjamin, Ph.D.

Experienced Forensic Toxicologist



- ◆ Analysis of Results of Blood, Urine, & Hair Drug Tests
- ◆ Cocaine/Narcotics Issues: Possession vs. Personal Use
- ◆ Dram Shop & Vehicular Homicide
- ◆ Medical & Law School Teaching Experience
- ◆ Excellent Communicator

References Available

617-969-1393

www.doctorbenjamin.com

medlaw@doctorbenjamin.com

rates. The same cannot be said of false positives, however, and that is what should concern us – the wrongly accused being arrested because of flawed “science.” The reason why so many people over .08 percent and .10 percent BAC show 4+ HGN clues is that so many people have 4+ HGN clues at .04 percent, .05 percent, and .06 percent BAC.

Researchers conducted three SFST validation studies between 1995 and 1998. Colorado was first in 1995; Florida and San Diego followed in 1997 and 1998, respectively. NHTSA says that the Colorado study was the first full field study that utilized law enforcement officers experienced in the use of SFSTs. Moreover, NHTSA claims that correct arrest decisions were made 93 percent of the time based on the three-test battery, which was substantially better than the initial study results.¹⁵

The Florida SFST field validation study sought to answer the question of whether SFSTs are valid and reliable indices of the presence of alcohol when used under “present day” traffic and law enforcement conditions. According to NHTSA, police officers made the correct arrest decisions 95 percent of the time.¹⁶ NHTSA goes on to say that the validation

studies have shown that the SFST three-test battery is the only scientifically validated and reliable method for discriminating between impaired and unimpaired drivers.

In undertaking the San Diego study, NHTSA wanted the SFSTs to be recognized as capable of discriminating BACs above and below .08 percent, as NHTSA and MADD campaigned to reduce per se limits to .08 percent across the nation. Not surprisingly, officers made the “correct arrest decision” 91 percent of the time at the .08 percent level and above.¹⁷

The Rest of the Story

The validation studies clearly demonstrate that HGN is generally present at BACs far below the per se limit of .08 percent. In the Colorado study, one in eight people under .05 had four or more HGN clues.¹⁸ In the Florida study, 18 percent of people below .08 percent BAC had five or six HGN clues.¹⁹ More significantly, it suggests that 50 percent of those under .08 percent had at least four clues, but it does not just come out and say it. NHTSA attempts to conceal these numbers by saying that half of the correctly released drivers had zero or two HGN clues. That

seems to say that half of the correctly released drivers (under .08 percent) had more than two clues.²⁰ Since an odd number of clues is highly unusual, a person with more than two clues probably has at least four. In yet another development that has not found its way into an SFST instructor or student manual, the Florida report also acknowledged that 67 percent of all incorrect arrests (under .08 percent) had all six clues.²¹

The other field sobriety tests also showed similar weaknesses. For example, the Florida study also found that those wrongly arrested averaged 3.6 clues on the walk-and-turn and two clues on the one-leg stand.²²

Another confounding factor in the field validation studies was the average breath alcohol concentration of those arrested. In the San Diego study, the one most frequently cited, the average BAC of those arrested was .15 percent.²³ The same was true in Florida.²⁴ In Colorado the average was a bit higher at .152 percent.²⁵ In other words, it should not generally be difficult to determine if someone is driving under the influence of alcohol when the BAC is this high, and the field sobriety tests have little influence on the decision to arrest as a practical matter. Nevertheless, the officers did have a great deal of difficulty at lower BACs, as evidenced by that fact that false positives (person arrested had BAC under .08 percent) were six times as common as a false negatives (person not arrested had BAC of .08 percent or more) in the San Diego study.²⁶ This last revelation has significant implications because all police officers participating in the San Diego study were equipped with NHTSA-approved portable breath testing devices.

The San Diego study yielded some other surprising results. Over one in three drivers (actually 30 out of 81, or 37 percent) under .08 percent had at least four HGN clues.²⁷ The physical dexterity FSTs fared even worse. Over one-half of those with a BAC under .08 percent (40 out of 76, or 53 percent) had two or more clues on the walk-and-turn.²⁸ In addition, over 40 percent of the people with a BAC under .08 percent (31 of 75, or 41 percent) had two or more clues on the one-leg stand.²⁹

The Florida results were even worse. At least 70 percent of everyone under .08 percent in the study demonstrated two or more clues on the walk-and-turn. This 70 percent figure only includes the *correctly released*; the actual percentage would be higher if the wrongly arrested were included.³⁰



NACDL STRIKE FORCE

YOU NEVER STAND ALONE

NACDL's Strike Force will review your case at no cost when you have been:

- Subpoenaed for properly representing a client
- Threatened with contempt
- Hit with an improper motion to disqualify you from a case

STRIKE FORCE CO-CHAIRS

Howard M. Srebnick
Miami, FL
(305) 371-6421
srebnick@royblack.com

Susan W. Van Dusen
Miami, FL
(305) 854-6449
svandusenlaw@aol.com

Martin G. Weinberg
Boston, MA
(617) 227-3700
owlmcb@worldnet.att.net

CIRCUIT COORDINATORS

For immediate assistance call the *Lawyers' Strike Force* Circuit Coordinator nearest you.

<p>1st Circuit Frank D. Inserni San Juan, PR (787) 763-3851 finserni@prtc.net</p> <p>Martin G. Weinberg Boston, MA (617) 227-3700 owlmcb@worldnet.att.net</p> <p>2nd Circuit William I. Aronwald White Plains, NY (914) 946-6565 waronwald@aol.com</p> <p>3rd Circuit Alan L. Zegas Chatham, NJ (973) 701-7080 alanatlaw@aol.com</p>	<p>4th Circuit Martin S. Pinales Cincinnati, OH (513) 721-4876 mpinales@cinci.rr.com</p> <p>John K. Zwerling Alexandria, VA (703) 684-8000 jz@zwerling.com</p> <p>5th Circuit Frank Jackson Dallas, TX (214) 871-1122 fjackson22@yahoo.com</p> <p>Kent A. Schaffer Houston, TX (713) 228-8500 zackymax@aol.com</p>	<p>6th Circuit James A. H. Bell Knoxville, TN (865) 637-2900 jbell@jamesahbell.com</p> <p>Donald A. Bosch Knoxville, TN (865) 637-2142 dbosch@boschlafirm.com</p> <p>7th Circuit Richard Kammen Indianapolis, IN (317) 236-0400 rkamm@iquest.net</p> <p>8th Circuit Ronald I. Meshbesh Minneapolis, MN (800) 274-1616 meshbesh@meshbesh.com</p>	<p>Burton H. Shostak St. Louis, MO (314) 725-3200 bshostak@shostaklawfirm.com</p> <p>9th Circuit Richard A. Cremer Roseburg, OR (541) 672-1955 rcremer@rosenet.net</p> <p>Alfred Donau, III Tucson, AZ (520) 795-8710 skipdonau@aol.com</p> <p>David A. Elden Los Angeles, CA (310) 478-3100 elden@innocent.com</p> <p>10th Circuit Charles W. Daniels Albuquerque, NM (505) 827-4889 charleswdaniels@gmail.com</p>	<p>Michael L. Stout Las Cruces, NM (505) 524-1471 mlstout@nm.net</p> <p>11th Circuit Howard M. Srebnick Miami, FL (305) 371-6421 srebnick@royblack.com</p> <p>Susan W. Van Dusen Miami, FL (305) 854-6449 svandusenlaw@aol.com</p> <p>DC Circuit Henry W. Asbill Washington, DC (202) 986-8141 hasbill@dl.com</p>
---	--	--	--	---

The data from the San Diego study show that the FSTs miserably fail the "specificity" test when it comes to those with a BAC between .07 percent and .09 percent. Specificity may be defined as the percentage of true negatives (i.e., under .08 percent) that are correctly classified as such by the test. The rate was only 36 percent.³¹ The rate was only slightly better at 44 percent for those between .06 percent and .10 percent, and increased to only 55 percent for those between .05 percent and .11 percent. Many people have some two-sided decision-making devices in their pockets, the most valuable of which is worth 25 cents, which would be just about as accurate.

The field validation studies uniformly report that the tests were performed correctly nearly every time, which is contrary to the experience of any experienced DUI attorney. For example, the Colorado study reported that only 13 errors of administration and six errors in instructions were observed in 305 SFST administrations, although only 41 percent were observed.³² More remarkably, no errors were observed in the 313 SFST batteries given in the Florida study, although only two-thirds of the administrations were monitored by *civilian employees of the same department*.³³ Simply stated, secretaries and other clerical personnel were grading the uniformed officers, and the report makes no mention of any training given to these civilian employees.

Although the standardized field sobriety tests may be of some limited utility, NHTSA and the police have clearly oversold their efficacy. Unfortunately, they had a head start of several years before the defense bar began to mount serious challenges to the tests, and by then a majority of the judiciary had tacitly (if not explicitly) accepted the field sobriety tests as an accurate means of determining impairment. The numbers from NHTSA's own field validation studies cast long shadows over the validity of the tests, and both the bench and the bar should take note. The bench should take a fresh look at the purposes for which the tests are admitted into evidence and reconsider how much weight they should be given in determining probable cause to arrest and proof warranting a conviction. Defense counsel should be ready to pounce whenever an officer is allowed to testify that the tests are 91 percent (or whatever number) accurate in determining if someone is under the influence. The statistics recited in this article demonstrate that NHTSA's claims are inflated and also provide ample material for an effective cross-examination.

Notes

1. M. Burns & H. Moskowitz, National Highway Traffic Safety Administration (NHTSA), *Psychophysical Tests for DWI Arrest*, Final Report, DOT-HS-802-424 (1977).
2. *Id.* at 25.
3. *Id.* at 8, Figure 4.
4. V. Tharp, M. Burns & H. Moskowitz, NHTSA, *Development and Field Test of Psychophysical Tests for DWI Arrest*, Final Report, DOT-HS-805-864 (1981).
5. *Id.* at 27, Table 8.
6. *Id.* at 22, Table 4.
7. *Id.* at 22, Table 4.
8. R.H. Nowaczyk & S. Cole, *Separating Myth From Fact: A Review of Research on the Field of Sobriety Tests*, THE CHAMPION, Aug. 1995, at 40.
9. V. Tharp, M. Burns & H. Moskowitz, *supra* at 15, Table 4.
10. M. Burns & H. Moskowitz, at 19, Figure 5.
11. V. Tharp, M. Burns & H. Moskowitz, *supra* at 35, Table 14.
12. *Id.* at 14, Table 2.
13. G. Good & C. Augsburger, *Use of Horizontal Gaze Nystagmus as a Part of Roadside Sobriety Testing*, 63 AM. J. OPTOMETRY & PHYSIOLOGICAL OPTICS 467-471 (1986).
14. *Id.* at 470, Table 2.
15. M. Burns & E. Anderson, NHTSA, *A Colorado Validation Study of the Standardized Field Sobriety Test (SFST) Battery*, Final Report, Project No. 95-408-17-5 (1995).
16. M. Burns & T. Dioquino, NHTSA, *A Florida Validation Study of the Standardized Field Sobriety Test (SFST) Battery*, Final Report, Project No. AL-97-05-14-01, Figure 5 (1997).
17. J. Stuster & M. Burns, NHTSA, *Validation of the Standardized Field Sobriety Test Battery at BACs Below 0.10 Percent*, Final Report, p. 18 (1998).
18. M. Burns & E. Anderson, *supra* at Figure 12.
19. M. Burns & T. Dioquino, *supra* at Table 4.
20. *Id.* at Section V, Subsection C, Topic 1.

21. *Id.*
22. M. Burns & T. Dioquino, *supra* at Table 4.
23. J. Stuster & M. Burns, *supra* at 16.
24. M. Burns & T. Dioquino, *supra* at Section V, Subsection A, Topic 1.
25. M. Burns & E. Anderson, *supra* at Section V, Subsection D.
26. J. Stuster & M. Burns, *supra* at 18, Figure 4.
27. *Id.* at 21, Figure 5.
28. *Id.*
29. *Id.*
30. M. Burns & T. Dioquino, *supra* at Section V, Subsection C, Topic 2.
31. M.P. Hlastala, N.L. Polissar & S. Oberman, *Statistical Evaluation of Standardized Field Sobriety Tests*, J. FORENSIC SCI., Vol. 50, No. 3, Table 3 (2005).
32. M. Burns & E. Anderson, *supra* at Section V, Subsection F.
33. M. Burns & T. Dioquino, *supra* at Section V, Subsection E. ■

About the Author

Allen Trapp, who currently serves as the National College for DUI Defense delegate from Georgia, was recently chosen to serve as a Georgia Association of Criminal Defense Lawyers vice president. He was named DUI Lawyer of the Year in 2005 by Georgia DODD (Defense of Drinking Drivers).



Allen M. Trapp Jr.
114-B Corporate Drive
P.O. Box 2206
Carrollton, GA 30117
770-830-8560
Fax 770-830-8565
E-MAIL amtjr@iscope.com

HOUSTON

NACDL's 1st Annual Defending Drug Cases Seminar

see brochure page 3

Register at www.nacdl.org/meetings
The Doubletree Hotel Downtown, Nov. 12-14, 2008