2014

# Standardized Field Sobriety Test: False Positive Test Rate among Sober Subjects

Keith Yoshizuka
*Touro University California*, keith.yoshizuka@tu.edu

Paul J. Perry
*Touro University California*, paul.perry@tu.edu

Greta Upton

Ingrid C. Lopes
*Touro University California*, ingrid.lopes@tu.edu

Eric Ip
*Touro University California*, eric.ip@tu.edu



## Recommended Citation

Yoshizuka, K., Perry, P. J., Upton, G., & Ip, E. J. (2014). Standardized field sobriety test: False positive test rate among sober subjects. *Journal of Forensic Toxicology and Pharmacology, 3*(2), [Article 120].

Yoshizuka et al., J Forensic Toxicol Pharmacol 2014, 3:2
http://dx.doi.org/10.4172/2325-9841.1000120

Journal of Forensic Toxicology & Pharmacology

Research Article

A SCITECHNOL JOURNAL

# Standardized Field Sobriety Test: False Positive Test Rate among Sober Subjects

**Keith Yoshizuka*, Paul J Perry, Greta Upton, Ingrid Lopes and Eric J Ip**

## Abstract

The Standardized Field Sobriety Test (SFST) is a series of exercises that a law enforcement officer gives to a driver suspected of driving under the influence of alcohol. The original research that demonstrated a high correlation between failure of the SFST and a high blood alcohol concentration did not utilize a standard control group to validate that the failure of the SFST was not a characteristic of the population at large. This study examined a series of drug naive subjects to determine the rate of failure of the SFST to accurately distinguish a suspect with high blood alcohol content from the general public. Of the 185 subjects tested, 26% of the drug naïve subjects failed the SFST. Since the SFST is used as evidence of probable cause to justify an arrest, a 26% false positive rate in the SFST may imply that the SFST may be only a minor factor in combination with other articulated evidence to justify sufficient probable cause for an arrest for driving under the influence, and may affect the weight of the evidence given to the SFST.

## Keywords

Standardized field sobriety test; False positives; Blood alcohol concentration

## Introduction

In 1975, the National Highway Traffic Safety Administration (NHTSA) endorsed research that correlated the Standardized Field Sobriety Test (SFST) with blood alcohol concentrations (BACs) [1,2]. In an effort to standardize and reliably assess impaired drivers, researchers evaluated tests that were most commonly used by officers at the time, including the one-leg stand (OLS), walk-and-turn (WAT), finger-to-nose, finger count, horizontal gaze nystagmus (HGN), tracing, and alternate tests (Romberg body sway, subtraction, counting backward, letter cancellation). All these tests were perturbed by alcohol. However, statistical analyses concluded that the combined score of three tests, the HGN and two divided attention task tests (OLS and WAT), were the "best test set" to assess sobriety. Using discriminant analysis, the combined scores of the three tests predicted that law enforcement officers could correctly classify 83% of test subjects as either sober or intoxicated, i.e., a BAC of 0.10% or greater [3]. A confirmatory study funded by the NHTSA supported the previous findings [4]. Ten officers administered the SFST in a laboratory to 297 drinking volunteers with BACs ranging from 0 to 0.18%. The officers were able to accurately categorize 81% of subjects on the basis of the SFST test results as to being a BAC <0.10% or a BAC ≥ 0.10%. The inter-rater and test-retest reliabilities kappa coefficient for the test battery ranged from a statistically acceptable 0.60 to 0.80. Burns reviewed ten studies assessing the validity of sobriety tests and concluded that a failed SFST was a reliable indicator regarding the presence of alcohol but not other CNS depressant drugs [5].

In 1981, the NHTSA implemented the SFST that enabled law enforcement officers to determine driver impairment and establish cause for arrest [6]. In October 2000, Congress passed and President Clinton signed into a law a bill that lowered the national standard for impaired driving to a BAC of ≥ 0.08%. States that did not enforce this federal provision by October 2003 would progressively lose federal highway funding, thus currently establishing a BAC of 0.08% as the legal per se limit in all states [7]. Before the passage of the new federal BAC limit, the NHTSA sponsored a study that established the creditability of the 0.08% limit. The study found that the SFST could accurately differentiate drivers above or below the more sensitive BAC ≥ 0.08% in 91% of subjects tested [1]. Data analyses among 297 suspected motorists found the SFST to be extremely accurate in discriminating between BACs above and below 0.08% threshold. The mean estimated and measured BACs of the 297 motorists tested were 0.117% and 0.122%, respectively. Further, analyses found the HGN test to be the most predictive of the three components of the SFST battery (r=0.65); a higher correlation was obtained when the results of all three tests were combined (r=0.69). However, a major limitation of this study was that only drivers suspected of being under the influence of alcohol were evaluated. Thus, the study did not include a control group to evaluate the SFST pass rate [1]. Therefore, it was not established whether a person could fail the SFST without any form of drug impairment or a BAC of 0.0%.

The objective of this manuscript is to report baseline (sober) failure rates from three studies that perturbed the SFST by drug-induced challenges to the study subjects [8-10].

## Methodology

The three experimental studies that collected baseline control data on sober test subjects included trazodone (sedating antidepressant) versus acetaminophen (over-the counter non-narcotic analgesic [8]; diphenhydramine (sedating antihistamine) versus fexofenadine (non-sedating antihistamine) [9]; and dextromethorphan (over-the counter opioid cough suppressant) versus docusate sodium (stool softener) [10].

Only researchers identified in the institutional review board proposals had full access to the data, and all researchers completed the National Institutes of Health human subjects training program. The study procedures were performed in accordance with the ethical standards of the Declaration of Helsinki, and the protocol was approved by the Touro University institutional review board. Sample size decisions for the three studies were based upon power analyses calculations using data from prior studies that demonstrated that diphenhydramine, trazodone, and dextromethorphan impaired psychomotor performance. A summary of the methodology of each of the studies is presented below.

***Corresponding author:** Dr. Keith Yoshizuka, PharmD, MBA, JD, Associate Professor, College of Pharmacy, Touro University, California, USA, Tel: 707-638-5992; Fax: 707-638-5953; E-mail: keith.yoshizuka@tu.edu

**Received:** January 08, 2014 **Accepted:** April 09, 2014 **Published:** April 16, 2014

**Table 1:** Baseline SFST[a] failure (pre-drug administration) rates for 3 studies.

| Study | | HGNb n (%) | WAT[c] n (%) | OLS[d] n (%) | Overall combined failure rate n (%) |
|---|---|---|---|---|---|
| Dextromethorphan vs. DSS Study | Dextromethorphan 30mg (n=40) | 0 (0%) | 6 (15.0%) | 2 (5.0%) | 22 (28%) |
| | Docusate Sodium 200mg (n=40) | 1 (2.5%) | 12 (30.0%) | 3 (7.5%) | |
| Diphenhydramine vs. Fexofenadine | Diphenhydramine (n=40) | 0 (0%) | 7 (17%) | 4 (10%) | 14 (23%) |
| | Fexofenadine (n=20) | 0 (0%) | 4 (20%) | 1 (5%) | |
| Trazodone vs. Acetaminophen | Trazodone 100mg (n=30) | 3 (10.0%) | 7 (23.3%) | 1 (3.3%) | 13 (29%) |
| | Acetaminophen 650mg (n=15) | 0 (0.0%) | 4 (26.7%) | 2 (13.3%) | |
| Overall Failure Rates | n=185 | 4 (2.2%) | 40 (22%) | 13 (7.0%) | 49 (26%) |

SFST: Standardized Field Sobriety Test; HGN: Horizontal Gaze Nystagmus; WAT: Walk-And-Turn; OLS: One-Leg Stand

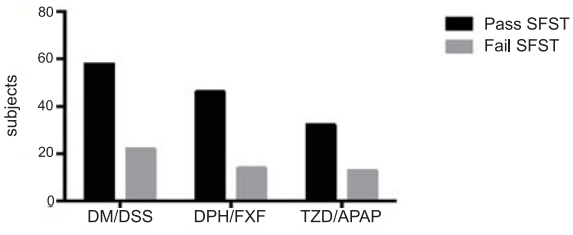


**Figure 1:** Baseline (sober) pass-fail rates for SFST for 3 studies (chi-square=0.48, p=0.79).

### Dextromethorphan versus docusate sodium (DM/DSS) [10]

This experiment was a randomized, double-blinded, repeated measures design involving 80 healthy adult participants. The study determined the failure rates on the SFST after a single dose of dextromethorphan 30mg or docusate sodium (DSS) 200mg. The SFST, administered by two physician-trained evaluators, was evaluated before drug ingestion (i.e. at baseline) and two hours post dextromethorphan or DSS ingestion. In this study, there was no statistical difference between the experimental and control test groups in subject age, gender, ethnicity, height, and weight or body mass index.

### Diphenhydramine versus fexofenadine (DPH/FXF) [9]

The goal of this experiment was to evaluate if the SFST could differentiate individuals who took diphenhydramine versus those who took fexofenadine. Using a randomized, double-blinded, repeated measures study design; 60 healthy adult individuals ingested either a single dose of diphenhydramine 50 mg or fexofenadine 60 mg. The SFST administered by two physician-trained evaluators, was evaluated before drug ingestion (i.e. at baseline) and two hours post dextromethorphan or fexofenadine ingestion.

### Trazodone versus acetaminophen (TZD/APAP) [8]

The goal of this study was to evaluate the passage of the SFST after a single dose of trazodone 100 mg or acetaminophen 650 mg. A randomized, double-blinded, repeated-measures design was employed. Forty-five healthy adult subjects were administered the SFST by two NHTSA-manual trained evaluators at baseline and two hours post trazodone (30 subjects) or acetaminophen (15 subjects) ingestion. SFSTs were conducted and evaluated by two trained individuals.

For the dextromethorphan and diphenhydramine studies, the SFST was administered at baseline and 2 hours post drug ingestion. SFST ratings were scored by two physician-trained evaluators. The physician, an experienced college professor, studied and then adapted the 161-page DWI Testing and Standardized Field Sobriety Testing student manual to train the evaluators [11]. Emphasis was placed on the mechanics of the SFST, the physiology and pathophysiology gauged by the test, and the nuances that enable testing to pass scrutiny in a court of law. Training videos assembled by law enforcement were reviewed multiple times. Each evaluator met with the physician in a workshop in order to practice each maneuver with both the verbal description and physical demonstration required by law. For the trazodone study, the raters were trained by two clinically experienced clinical pharmacists using the same DWI Testing and Standardized Sobriety Field Testing student manual.

## Results

Table 1 presents a summary of the baseline failure rates for the studies. Importantly, the SFST is scored as failed if the test subject fails either the HGN, WAT or OLS. In the dextromethorphan versus DSS study, 1 subject failed the HGN, 18 subjects failed the WAT, and 5 subjects failed the OLS. Overall, 22 of 80 (28%) subjects failed the SFST. In the diphenhydramine versus fexofenadine study, 11 failed the WAT and 5 failed the OLS, resulting in an overall failure rate of 14 of 6o subjects (23%). Finally, in the trazodone versus acetaminophen study, 3 subjects failed the HGN, 11 failed the WAT, and 3 failed the OLS. Overall 13 out of 45 (29%) of the subjects failed. Considering the total control and intervention study populations from the three studies, 49 out of 185 subjects failed some aspect of the SFST at baseline. In other words, there was a 26% failure rate of the SFST when no form of pharmacologic perturbation to the study subjects occurred. Figure 1 presents the pass-fail rates for the three independent studies. A chi-square test of independence was performed on these data to determine if there were differences pass-fail frequency (count) data between the 3 studies. The p-value of 0.46 indicates that the pass-fail rates between the studies did not differ. Thus the SFST failure rate was reproducible between different groups of subjects.

## Discussion

The SFST is commonly used throughout the country to test for impairment while driving. The original research demonstrated a high correlation between failure of the SFST and a BAC of ≥ 0.08%. The limitation of the Stuster and Burns [1] data is that there was no control group to compare the results. However, this deficiency can be remedied by using the data of Stuster and Burns [1] as a historical experimental group. The authors acknowledge that use of such a control group is a limitation as it may introduce a bias. Thus, a 2×2

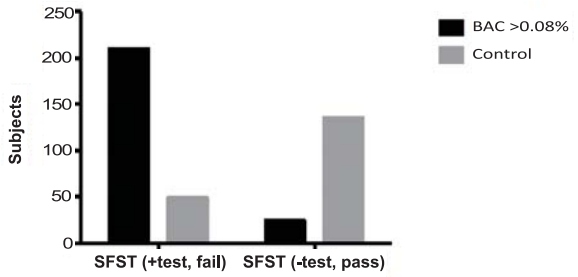


**Figure 2:** SFST pass-fail values for 234 (fail=210) subjects with BAC> 0.08% versus 185 (fail=49) drug-naive subjects (p, 0.0001).

contingency table was constructed. The control group consisted of the present study's data of 185 subjects of whom 49 who had a positive or failed SFST. The alcohol consumer group consisted of the Stuster and Burns 234 subjects who tested positive for a BAC of >0.08% BAC of whom 210 had a positive or failed SFST. As illustrated in Figure 2, a significant difference was obvious between the two groups (Fisher's exact test p<0.0001, sensitivity=0.90, specificity=0.74, positive predictive value 0.81 and negative predictive value=0.85). More importantly, the data demonstrate that there is a false positive rate of 0.26 and false negative rate of only 0.10 that is problematic from a legal standpoint. Additionally, examination of the entire Stuster and Burns [1] data set of 297 subjects finds that of 83 individuals with BAC <0.08%, 24 (29%) were characterized by the officers as having a BAC of ≥ 0.08%. Stated in another way, their false positive rates are nearly identical to our sober or drug naïve subject 26% SFST failure rate. A control group is used to establish that the correlation has a meaningful relationship to the outcome, and is not just a characteristic of the population at large.

Harris identified 23 viable visual clues that are used by police officers to detect drivers driving under the influence [12]. He generated a Drunk Driver Detection Guide as an aid for use by officers. He found that there was an association between the number of clues and the probability of the driver having a BAC of either >0.1% or >0.05%. As an example, if a nighttime driver had been noted to be following a car too closely, the probability of a ≥0.1% BAC was 55%. If two additional clues of straddling the center-line and slow response to traffic signals were observed, the probability of a DUI increased to 65%. The same three clues generated an 85% probability of a ≥0.05% BAC. The Harris system accuracy nearly approaches the SFST 91% accuracy rate reported by Stuster and Burns [1].

The observation that visible clues of Harris [12] and the SFST data of Stuster and Burns [1] generate similar accuracy rates provokes the question as to whether some component(s) of the SFST have a negative effect on the accuracy rates. Inspection of Table 1 notes a HGN failure rate of only 2.2% versus 21.6% and 7.0% failure rates for the WAT and OLS respectively. Stuster and Burns had only a 1.9% (4/209) failure rate on the HGN [1]. They have referred to the HGN as contributory evidence that "provides indisputable evidence of alcohol in a motorist's system" [1]. The explanation for this "dogma" is that experienced drinkers can perform the voluntary physical divided attention task tests, i.e., OLS and WAT correctly even with a BAC >0.01% but they cannot pass the HGN because it is an involuntary reaction over which they have no voluntary control.

With a false positive rate of 26%, the authors submit that the SFST can only be contributing evidence to justify an arrest for driving under the influence, in combination with other articulated evidence such as the officer's observation of erratic driving behavior, the odor of alcohol on the driver's breath, or other such evidence in order to have sufficient probable cause to justify an arrest.

What this study demonstrates is that there are also a significant number of persons NOT impaired who cannot successfully pass the SFST. The SFST is used by law enforcement as evidence of probable cause to justify an arrest of the subject. The results of this study call into question the validity of using the SFST as the primary justification for an arrest for driving under the influence. Other articulated evidence to justify the initial traffic stop such as an obvious equipment violation, a moving violation, unusual or suspicious behavior, or almost anything else that would call attention to the suspect vehicle would be required [11]. Once the vehicle is stopped, the officer must observe and interview the driver face-to-face. At this point, the officer must be able to articulate further evidence to justify requiring the driver to exit the vehicle to administer the SFST. Probable cause is a level of reasonable belief, based on articulated fact, required to arrest and prosecute a person in criminal court that a reasonable person would find sufficient for a conviction. The quantum of evidence required for arrest is generally a preponderance of the evidence, as compared to a conviction that requires a quantum of beyond a reasonable doubt. This study provides information that affects the weight given to the SFST as evidence. Only when all of these conditions are met is it reasonable to administer the SFST, and even failing the SFST is not dispositive of intoxication absent further evidence. Officers do not use any single test as a basis to justify arrest, but must evaluate the totality of the evidence of exhibited behaviors, performance tests, and other observed evidence as the basis for arrest.

## References

1. Stuster J, Burns M (1998) Validation of the standardized field sobriety test battery at BACs below 0.10 percent, U.S. Department of Transportation, National Highway Traffic Safety Administration.
2. Stuster J (2006) Validation of the standardized field sobriety test battery at 0.08% blood alcohol concentration. Hum Factors 48: 608-614.
3. Burns M, Moskowitz H (1977) Psychophysical Tests for DWI Arrest, U.S. Department Of Transportation, National Highway Traffic Safety Administration, Washington.
4. Tharp V, Burns M, Moskowitz H (1981) Development and field test of psychophysical tests for DWI arrest, U.S. Department Of Transportation, National Highway Traffic Safety Administration, Washington.
5. Burns M (2003) An overview of field sobriety test research. Percept Mot Skills 97:1187-1199.
6. Development of a standardized field sobriety test (2001) U.S. Department Of Transportation, National Highway Traffic Safety Administration, Washington.
7. U.S. Department of Transportation, National Highway Traffic Safety Administration, Federal Highway Administration (2003) 68 FR 6091-Operation of motor vehicles by intoxicated persons.
8. Ip EJ, Bui QV, Barnett MJ, Kazani A, Wright R, et al. (2013) The effect of trazodone on standardized field sobriety tests. Pharmacotherapy 33: 369-374.
9. Lu D, Ip E, Lopes I, Barnett M, Chu E, et al. (2014) Effects of diphenhydramine versus fexofenadine on the standardized field sobriety test. 14th Annual Research Day, Touro University-California, March, Vallejo, CA.
10. Pal J, Ip E, Trinh K, Yu J, Lindfelt T, et al. (2013). Effects of dextromethorphan on the standardized field sobriety test. 13th Annual Research Day, Touro University-California, March, Vallejo, CA.
11. U.S. Department of Transportation (2006) DWI (driving while intoxicated) Detection and Standardized Field Sobriety Testing.
12. Harris DH (1980) Visual detection of driving while intoxicated. Hum Factors 22: 725-32.

## Author Affiliations

Top

*College of Pharmacy Touro University, California, USA*