Enhancing the Standardized Field Sobriety Test to Detect Cannabis Impairment

Douglas J. Beirness, Ph.D., Senior Research Associate, Canadian Centre on Substance Use and Addiction
Amy J. Porath, Ph.D., Director of Research, Canadian Centre on Substance Use and Addiction
D’Arcy Smith, OIC Drug Evaluation and Classification Program, Royal Canadian Mounted Police
Disclosures

Relationships with commercial interests:
None

Potential for conflict(s) of interest:
None

Funding: Ontario Ministry of Transportation
AAA Foundation for Traffic Safety
Canadian Centre on Substance Use and Addiction
Overview

- Standardized Field Sobriety Test (SFST)
- Validation of SFST
  - Reliability
  - Validity
- SFST for Drug Impairment
- New tests/indicators
- Enhancements
Components of the SFST

- One Leg Stand
- Horizontal Gaze Nystagmus
- Walk and Turn
SFST

- **Horizontal Gaze Nystagmus Test (HGN)**
  - Eyes follow a slowly moving object as it is moved from side to side

- **One Leg Stand Test (OLS)**
  - Stand with one foot about 15 cm off the ground while counting aloud for 30 seconds

- **Walk and Turn Test (WAT)**
  - Walk nine steps, heel-to-toe, along a straight line, followed by a turn on one foot, and then return in the same manner in the opposite direction
SFST Clues

- Horizontal Gaze Nystagmus Test (HGN)
  1. Lack of smooth pursuit
  2. Nystagmus at maximum deviation
  3. Nystagmus prior to 45°

- One Leg Stand Test (OLS)
  1. Sways while balancing
  2. Raises arms
  3. Hops
  4. Puts foot down
SFST Clues

- Walk and Turn Test (WAT)
  1. Loses balance during instructions
  2. Starts too soon
  3. Stops while walking
  4. Misses heel to toe
  5. Steps off line
  6. Raise arms
  7. Wrong number of steps
  8. Improper turn
SFST Background

- Burns and Moskowitz (1977)
- Search for a test police could use at roadside to identify drivers who were impaired by alcohol
- Began with series of potential tests from other studies
- Administered different doses of alcohol
- Light, moderate and heavy drinkers
- Three of the best tests that were deemed easy to use at roadside were selected
SFST Studies

- Burns & Moskowitz (1977)
- Tharp, Burns & Moskowitz (1981)
- Burns & Anderson (1995) (Colorado)
- Stuster (1997)
- Stuster & Burns (1998) (San Diego)
- Burns & Dioguino (1998) (Florida)
What Problem Are We Trying to Solve?

- The SFST was initially developed and validated to detect drivers impaired by alcohol.
- The test is currently used by police to detect drug impairment.
- SFST’s ability to accurately detect drug-related impairment is unclear.
- There have been few studies that have assessed the SFST’s ability to accurately detect impairment due to drugs.
  - Controlled studies in laboratory settings.
Reliability is the extent to which a measuring instrument, device, or test provides consistency in measurement.
Reliability

- Reliability is concerned with consistency in measurement.
- Different types of Reliability:
  1. Test-retest reliability
  2. Inter-rater reliability
Reliability: Test-Retest

Tested same subjects on two separate occasions at same BAC by same officer.

- HGN 0.66
- WAT 0.72
- OLS 0.61
- Total 0.71

(Tharp, Burns & Moskowitz, 1981)
Reliability: Inter-rater

- **Data Collection**
  - DRE certification sessions
  - 2 evaluators scoring at the same time
  - 248 paired observations
  - data used assess inter-rater reliability
Reliability  
Walk and Turn

<table>
<thead>
<tr>
<th>Clue</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance</td>
<td>87%</td>
</tr>
<tr>
<td>Too Soon</td>
<td>92%</td>
</tr>
<tr>
<td>Stops</td>
<td>75%</td>
</tr>
<tr>
<td>Miss H/T</td>
<td>72%</td>
</tr>
<tr>
<td>Off Line</td>
<td>79%</td>
</tr>
<tr>
<td>Raise Arms</td>
<td>81%</td>
</tr>
<tr>
<td># Steps</td>
<td>91%</td>
</tr>
<tr>
<td>Turn</td>
<td>81%</td>
</tr>
<tr>
<td>2+ Clues</td>
<td>87%</td>
</tr>
<tr>
<td>Clue</td>
<td>Agreement</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Sway</td>
<td>76%</td>
</tr>
<tr>
<td>Arms</td>
<td>80%</td>
</tr>
<tr>
<td>Hops</td>
<td>94%</td>
</tr>
<tr>
<td>Foot down</td>
<td>90%</td>
</tr>
<tr>
<td># Clues</td>
<td>78%</td>
</tr>
<tr>
<td>2+ Clues</td>
<td>87%</td>
</tr>
</tbody>
</table>
Reliability
Horizontal Gaze Nystagmus

Clue               Agreement
----------         ------------
LSP               96% (n=33)
Max Dev           99% (n=21)
Onset <45°        99% (n=19)

(most subjects did not show HGN)
Validity

Validity is the extent to which a measuring instrument, device, or test measures what it is supposed to measure.
Validity

Types of Validity

- **Construct** validity – What is it supposed to measure?
- **Content** validity – Does it adequately sample from the construct you are trying to measure?
- **Face** validity – Does it look like it measures what it is supposed to?
- **Criterion** validity – Is it predictive of some criterion measure?
Validity -- Intelligence

- **Construct** – Does the test measure intelligence?
- **Content** – Does the test adequately sample from the domain of “intelligence”?
- **Face** – Does it look like a test of intelligence?
- **Criterion** – Is it predictive of things we would expect of intelligent people?
Decision Matrix

<table>
<thead>
<tr>
<th>Test</th>
<th>Criterion (&quot;Truth&quot;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Negative: True Negative</td>
</tr>
<tr>
<td>Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Test</td>
<td>Negative</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>Negative</td>
<td>True Negative</td>
</tr>
<tr>
<td>Positive</td>
<td>False Positive</td>
</tr>
</tbody>
</table>

**Criterion (“Truth”)**

- **Sensitivity** = proportion of all positive cases correctly identified by the Test -- $TP/(TP+FN)$
- **Specificity** = proportion of all negative cases correctly identified by the test -- $TN/(TN+FP)$
- **Miss Rate** = proportion of all positive cases incorrectly identified by the test as negative -- $FN/(TP+FN)$
- **False Positive Rate** = proportion of all negative cases incorrectly identified by the test as positive -- $FP/(TN+FP)$
- **Accuracy** = proportion of all cases correctly classified – $TN+TP/Total$ cases
### Decision Matrix: Prostate Cancer

<table>
<thead>
<tr>
<th>DRE + PSA</th>
<th>Biopsy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative</td>
<td>Positive</td>
</tr>
<tr>
<td>Negative</td>
<td>True Negative</td>
<td>False Negative</td>
</tr>
<tr>
<td>Positive</td>
<td>False Positive</td>
<td>True Positive</td>
</tr>
</tbody>
</table>

- **Sensitivity** = .205
- **Specificity** = .936
- **Miss Rate** = .795
- **False Positive Rate** = .064
Validation of the SFST

- **Construct Validity** - What is Impairment?
  - Change from alcohol/drug-free baseline
  - Reduced ability to operate vehicle safely

- **Content Validity**
  - Multitude of tests used in experimental literature
  - Do tests measure impairment of driving skills?

- **Face Validity**
  - Balance, coordination, follow instructions

- **Predictive/Criterion Validity**
  - Is test related to other measures of impairment?
What’s the Criterion?

- Correlation between amount of alcohol (BAC) and degree of impairment

- Over time, operational definition of “impairment” has become “BAC” – i.e., if BAC > .08, driver is deemed impaired

- Criterion for impairment is BAC ≥ .08

- Compare with SFST performance
# Decision Matrix

<table>
<thead>
<tr>
<th>SFST -- Officer Decision</th>
<th>BAC &lt; .08</th>
<th>BAC ≥ .08</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Impaired (Release)</td>
<td>True Negatives</td>
<td>False Negatives</td>
</tr>
<tr>
<td>Impaired (Arrest)</td>
<td>False Positives</td>
<td>True Positives</td>
</tr>
</tbody>
</table>
**Stuter & Burns (1998)**

### Overall Accuracy
- **All 3** = 91%
- **HGN** = 88%
- **WAT** = 79%
- **OLS** = 83%

<table>
<thead>
<tr>
<th>Decision</th>
<th>BAC &lt; .08</th>
<th>BAC ≥ .08</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Impaired</td>
<td>True Negatives 59</td>
<td>False Negatives 4</td>
</tr>
<tr>
<td>Impaired</td>
<td>False Positives 24</td>
<td>True Positives 210</td>
</tr>
</tbody>
</table>
Sensitivity

Of all cases > .08, in how many did the SFST predict correctly?

- HGN = 98%
- WAT = 92%
- OLS = 92%
- All 3 = 98%

<table>
<thead>
<tr>
<th>Decision</th>
<th>Criterion</th>
<th>BAC &lt; .08</th>
<th>BAC ≥ .08</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Impaired</td>
<td>True Negatives n=59</td>
<td>False Negatives n=54</td>
<td></td>
</tr>
<tr>
<td>Impaired</td>
<td>False Positives n=24</td>
<td>True Positives n=210</td>
<td></td>
</tr>
</tbody>
</table>
Stuter & Burns (1998)

**Specificity**

Of all cases < .08, in how many did the SFST predict correctly?

- HGN = 63%
- WAT = 47%
- OLS = 59%
- All 3 = 71%

False Positive Rate = 29%
Validity of the SFST

- Has the elements required for a valid test of impairment due to alcohol
- Has good criterion/predictive validity – i.e., accurately detects drivers with BACs of at least .08
- Each component of the test (HGN, WAT, OLS) shows good performance statistics
- Watch for “false positives”
Is the SFST a valid test to identify driver impairment due to drugs?
Study Sample

- Sample of 2,142 Drug Evaluation and Classification case reports
  - Central Nervous System (CNS) stimulants ($n = 852$)
  - Cannabis ($n = 703$)
  - Narcotic analgesics ($n = 312$)
  - CNS depressants ($n = 135$)
  - No drug cases ($n = 140$)

- The opinion of the Drug Recognition Expert matched the toxicological analysis of the bodily fluid sample
Overall Results

- Multinominal logistic regression conducted to predict drug category
- Together, all four drug categories were significantly associated with impaired performance on the HGN, WAT and OLS tests of the SFST
- All significant at $p < .0001$
## Results: HGN Test

<table>
<thead>
<tr>
<th>Sign of Impairment</th>
<th>Affected Drug Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased odds of lack of smooth pursuit*</td>
<td>CNS depressants</td>
</tr>
<tr>
<td>Increased odds of distinct nystagmus at maximum deviation*</td>
<td>CNS depressants</td>
</tr>
</tbody>
</table>

*p < 0.0167.
Results: WAT Test

<table>
<thead>
<tr>
<th>Sign of Impairment</th>
<th>Affected Drug Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decreased odds of keeping balance while listening to test instructions*</td>
<td>CNS depressants</td>
</tr>
<tr>
<td></td>
<td>CNS stimulants</td>
</tr>
<tr>
<td></td>
<td>Narcotic analgesics</td>
</tr>
<tr>
<td>Decreased odds of touching heel-to-toe while walking*</td>
<td>CNS depressants</td>
</tr>
<tr>
<td>Decreased odds of taking the correct number of steps*</td>
<td>Narcotic analgesics</td>
</tr>
</tbody>
</table>

*<.0071
### Results: OLS Test

<table>
<thead>
<tr>
<th>Sign of Impairment</th>
<th>Affected Drug Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased odds of swaying while balancing on one leg*</td>
<td>All categories</td>
</tr>
<tr>
<td>Increased odds of using arms to maintain balance*</td>
<td>All categories</td>
</tr>
<tr>
<td>Increased odds of putting raised foot down*</td>
<td>CNS depressants</td>
</tr>
<tr>
<td></td>
<td>CNS stimulants</td>
</tr>
<tr>
<td></td>
<td>Narcotic analgesics</td>
</tr>
<tr>
<td>Decreased odds of hopping during test to maintain balance*</td>
<td>All categories</td>
</tr>
</tbody>
</table>

*p < .0125
Conclusions

- All drug categories were significantly associated with impaired performance on the SFST.
- The pattern of signs on the various tests of the SFST varied by drug category, providing support for the validity of using this test to detect persons impaired by drugs other than alcohol.
- Further research is needed using larger samples of drug-impaired drivers with known drug-blood concentrations.
- Normative data on individuals’ performance on the SFST while not under the influence of drugs is also needed.
Predictive/Criterion Validity

- Data from DRE evaluations provide a wealth of information that can be used to help determine validity of SFST to detect drug impairment
- With legalization, cannabis became the focus
- Used the decision matrix approach similar to that used to validate SFST for alcohol
What’s the Criterion?

- For alcohol, used BAC ≥ .08
- Correlation between amount of drug and degree of impairment
- Per se limits for drugs?
- Examples for cannabis (THC)
  - 5 ng/mL (e.g., WA, CO, Canada)
  - 2 ng/mL (NV, OH, UK, Canada)
  - Zero (11 states)
Study Sample 2

- Sample of 579 DEC evaluations with blood THC concentrations – no other substances
- The opinion of the Drug Recognition Expert matched the toxicological results
- Sample of 321 drug-free evaluations
- Start by looking at SFST clues
HGN Clues: Cannabis

Percent

LSP  | Max Dev  | Prior 45
---|---------|--------
<5 ng/ml | <5 ng/ml | <5 ng/ml
5+ ng/ml  | 5+ ng/ml  | 5+ ng/ml
One Leg Stand Clues: Cannabis

- Sway
- Arms
- Hops
- Foot Down

Percent

- <5 ng/ml
- 5+ ng/ml
Walk and Turn Clues: Cannabis

Percent

- Too Soon Lose Bal
- Stops
- Heel/Toe
- Off Line
- Arms
- Steps
- Turn

<5 ng/ml
5+ ng/ml
### Predictive/Criterion Validity

- Decision Matrix similar to original SFST validation studies for alcohol

#### Table: SFST Test Score vs. THC Concentration

<table>
<thead>
<tr>
<th>SFST Test Score</th>
<th>THC &lt; 5 ng/ml</th>
<th>THC 5+ ng/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Impaired</td>
<td>True Negatives</td>
<td>False Negatives</td>
</tr>
<tr>
<td>Impaired</td>
<td>False Positives</td>
<td>True Positives</td>
</tr>
</tbody>
</table>
## HGN: Cannabis

### Criterion

<table>
<thead>
<tr>
<th>HGN Score</th>
<th>THC &lt; 5 ng/ml</th>
<th>THC 5+ ng/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;4 Clues (Not Impaired)</td>
<td>374 (True Negatives)</td>
<td>431 (False Negatives)</td>
</tr>
<tr>
<td>4+ Clues (Impaired)</td>
<td>54 (False Positives)</td>
<td>31 (True Positives)</td>
</tr>
</tbody>
</table>

- Sensitivity = 7%
- Specificity = 87%
- Accuracy = 46%
## Walk and Turn: Cannabis

<table>
<thead>
<tr>
<th>WAT Score</th>
<th>THC &lt; 5 ng/ml</th>
<th>THC 5+ ng/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;2 Clues (Not Impaired)</td>
<td>91 (True Negatives)</td>
<td>103 (False Negatives)</td>
</tr>
<tr>
<td>2+ Clues (Impaired)</td>
<td>326 (False Positives)</td>
<td>351 (True Positives)</td>
</tr>
</tbody>
</table>

- **Sensitivity** = 77%
- **Specificity** = 22%
- **Accuracy** = 51%
Validity of SFST for Cannabis

- SFST unable to distinguish between THC concentration above and below 5 ng/ml
- HGN not detecting positive cases accurately
- WAT & OLS many false positives

*Is it the test or the criterion?*
HGN Clues

Percent

THC = 0
THC > 0

LSP
Max Dev
Prior 45
One Leg Stand Clues

![Graph showing the percent of THC = 0 and THC > 0 for different signs of impaired balance: Sway, Arms, Hops, and Foot Down.](graph.png)
Walk and Turn Clues

- THC = 0
- THC > 0

Bar chart showing the percent of each clue for THC = 0 and THC > 0.
<table>
<thead>
<tr>
<th>HGN Score</th>
<th>THC = 0</th>
<th>THC &gt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;4 Clues (Not Impaired)</td>
<td>407 (True Negatives)</td>
<td>1403 (False Negatives)</td>
</tr>
<tr>
<td>4+ Clues (Impaired)</td>
<td>28 (False Positives)</td>
<td>160 (True Positives)</td>
</tr>
</tbody>
</table>

- Sensitivity = 10%
- Specificity = 93%
- Accuracy = 28%
## Walk and Turn

### Criterion

<table>
<thead>
<tr>
<th>WAT Score</th>
<th>THC = 0</th>
<th>THC &gt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;2 Clues (Not Impaired)</td>
<td>277 (True Negatives)</td>
<td>276 (False Negatives)</td>
</tr>
<tr>
<td>2+ Clues (Impaired)</td>
<td>147 (False Positives)</td>
<td>1229 (True Positives)</td>
</tr>
</tbody>
</table>

- Sensitivity = 82%
- Specificity = 65%
- Accuracy = 78%
Study Sample 3

- Sample of 5,218 DEC evaluations with toxicology
- 1,608 THC only cases
- The opinion of the Drug Recognition Expert matched the toxicological results
- Sample of 437 drug-free evaluations
- Start by looking at SFST clues
WAT (2+ Clues) by Drug Category

- No Drug: 34.3
- Depressant: 98.1
- Stimulants: 91.9
- Narcotic Analgesic: 94.2
- Cannabis: 81.7
OLS (4+ Clues) by Drug Category

No Drug: 14.7
Depressant: 78.4
Stimulants: 62.4
Narcotic Analgesic: 71.1
Cannabis: 40.3
### Validity Indicators for SFST

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Drugs</td>
<td>0.607</td>
<td>0.867</td>
<td>0.634</td>
</tr>
<tr>
<td>Depressants</td>
<td>0.961</td>
<td>0.867</td>
<td>0.913</td>
</tr>
<tr>
<td>Stimulants</td>
<td>0.629</td>
<td>0.867</td>
<td>0.724</td>
</tr>
<tr>
<td>Narcotic Analgesics</td>
<td>0.698</td>
<td>0.867</td>
<td>0.793</td>
</tr>
<tr>
<td>Cannabis</td>
<td>0.414</td>
<td>0.867</td>
<td>0.513</td>
</tr>
</tbody>
</table>
Additional Tests/Indicators for Drugs

- Romberg Balance
- Finger to Nose
- Finger to Finger
- Finger Count
- Hand Pat
- Coin pick-up
- Head Movement/Jerks
- Lack of Convergence
- Eyelid Tremors
- Backwards Alphabet
Other Indicators
## Other Cannabis Indicators

<table>
<thead>
<tr>
<th>Test/Indicator</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC</td>
<td>63%</td>
<td>61%</td>
<td>62%</td>
</tr>
<tr>
<td>Romberg 2+</td>
<td>52%</td>
<td>79%</td>
<td>58%</td>
</tr>
<tr>
<td>Eyelid Tremors</td>
<td>78%</td>
<td>69%</td>
<td>77%</td>
</tr>
<tr>
<td>FTN 3+</td>
<td>83%</td>
<td>52%</td>
<td>76%</td>
</tr>
<tr>
<td>Droopy Eyelids</td>
<td>41%</td>
<td>86%</td>
<td>51%</td>
</tr>
</tbody>
</table>
SFST + FTN + Eyelid Tremors

<table>
<thead>
<tr>
<th>Category</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Drugs</td>
<td>0.949</td>
<td>0.681</td>
<td>0.915</td>
</tr>
<tr>
<td>Depressants</td>
<td>0.992</td>
<td>0.681</td>
<td>0.842</td>
</tr>
<tr>
<td>Stimulants</td>
<td>0.943</td>
<td>0.681</td>
<td>0.783</td>
</tr>
<tr>
<td>Narcotic Analgesics</td>
<td>0.949</td>
<td>0.681</td>
<td>0.782</td>
</tr>
<tr>
<td>Cannabis</td>
<td>0.937</td>
<td>0.681</td>
<td>0.879</td>
</tr>
</tbody>
</table>
Conclusions

- There is evidence to support the reliability and validity of SFST to detect cannabis impairment in drivers.
- HGN not an indicator of cannabis.
- May want to consider additional tests/indicators.
- Work is ongoing…
New Study...

- Enhancing the SFST with a supplemental test(s)
- Develop a scoring system for Finger to Nose (FTN) test
- Develop new procedure for Modified Romberg Balance (MRB)
- Controlled administration of cannabis – double blind placebo
- Use DREs to administer tests
- Determine best supplemental test(s) and scoring schemes to assess cannabis impairment
Save the Date!

Evidence and Perspectives, Compassion and Action.

- CCSA’s Issues of Substance is Canada’s premiere conference for the substance use and addiction field
- Registration opens March 2019
- #CCSAConference