Validating the Standardized Field Sobriety Test For Drugs

Doug Beirness
Ottawa, ON Canada

DRE Conference
Denver, August 2016
Overview

- Validation of SFST
- SFST for Drug Impairment
  - Validity
  - Reliability
- New tests/indicators?
- International practices
What Problem are we trying to solve?

- SFST developed to detect *alcohol* impairment and we need evidence to support the *validity* and *reliability* of the SFST to assess impairment by *drugs*
Validity

Validity is the extent to which a measuring instrument, device, or test measures what it is supposed to measure.
Reliability is the extent to which a measuring instrument, device, or test provides consistency in measurement.
Validity

Some key components of Validity:

- **Construct** validity – What is it supposed to measure?
- **Content** validity – Does it adequately sample from the domain you are trying to measure?
- **Face** validity – Does it look like it measures what it is supposed to?
- **Criterion** validity – Is it predictive of, or related to, some criterion measure?
Validity – Tire Pressure Gauge

- **Construct** validity – air pressure
- **Content** validity – force of air on wall of a tire
- **Face** validity
- **Criterion** validity
  - too much or too little
Validity -- Intelligence

- **Construct** – what is intelligence?
- **Content** – does 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?
Validation of the SFST

- **Construct Validity** - What is Impairment?
  - Change from alcohol-free baseline
  - Reduced ability to operate vehicle safely
  - SFST performance related to driving performance?

- **Content Validity**
  - Multitude of tests used in experimental literature
  - Do tests measure critical components of driving?

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

- **Predictive/Criterion Validity**
  - Is test predictive of driving impairment?
What’s the Criterion?

- Correlation between amount of alcohol and degree of impairment
- Over time, operational definition of “impairment” has become “BAC” – i.e., if BAC > .08, driver is impaired
- Officer arrest decision based on SFST compared to BAC
- Criterion for impairment is BAC ≥ .08
SFST Validation Studies

- Burns & Moskowitz (1977)
- Tharp, Burns & Moskowitz (1981)
- Burns & Anderson (1995) (Colorado)
- Stuster (1997)
- Stuster & Burns (1998) (San Diego)
- Burns & Dioguino (1998) (Florida)
Criterion Validity

- Using SFST, do officers make the correct arrest/release decision regarding suspected impaired drivers?

Burns & Moskowitz

- 6 tests, rated 0-10
- “Arrest” decision accuracy 83%
- Officers appeared “conservative” in arrest decisions (Miss rate=31%)
- (73% of cases had measured BAC < .10)
## Decision Matrix

<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</td>
<td></td>
<td>False Negatives</td>
</tr>
<tr>
<td>Impaired</td>
<td>False Positives</td>
<td>True Positives</td>
<td></td>
</tr>
</tbody>
</table>
Overall Accuracy
- All 3 = 91%
- HGN = 88%
- WAT = 79%
- OLS = 83%
Stuster & Burns (1998)

**Sensitivity**
Of all cases >.08, in how many did the officer get it right?
- 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</td>
<td></td>
<td>False Negatives</td>
</tr>
<tr>
<td>Impaired</td>
<td>False Positives</td>
<td>True Positives</td>
<td></td>
</tr>
</tbody>
</table>
Stuster & Burns (1998)

**Specificity**
Of all cases <.08, in how many did the officer get it right?
- HGN = 63%
- WAT = 47%
- OLS = 59%
- All 3 = 71%

<table>
<thead>
<tr>
<th>Decision</th>
<th>Criterion</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>BAC &lt; .08</strong></td>
<td><strong>BAC ≥ .08</strong></td>
<td></td>
</tr>
<tr>
<td>Not Impaired</td>
<td>True Negatives</td>
<td>False Negatives</td>
<td></td>
</tr>
<tr>
<td>Impaired</td>
<td>False Positives</td>
<td>True Positives</td>
<td></td>
</tr>
</tbody>
</table>

High rate of False Positives
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
Reliability is the extent to which a measuring instrument, device, or test provides consistency in measurement.
Reliability

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

Tharp et al. (1981)

2. Test-retest reliability

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
SFST and Alcohol

- SFST is a valid test for identifying driver impairment due to alcohol.
- Evidence for reliability of SFST is limited but results are acceptable.

Is the SFST a valid and reliable test to identify driver impairment due to drugs?
Validity – SFST for Drugs

- **Construct** – What is impairment?
- **Content** – Does test adequately sample from the domain of “impairment”?
- Does it look like a test of impairment?
- Is it predictive of some criterion?
Drug “Per Se” Limits

- Use drug per se limits as criterion
- Science isn’t strong on drug per se limits
- Examples for cannabis (THC)
  - 5 ng/ml (WA, CO, MT)
  - 2 ng/ml (NV, OH) (UK) (Norway 1.3 ng/ml)
  - 0.4 ng/ml (PA)
  - Zero (10 states)
Predictive/Criterion Validity

- Data from DRE evaluations provide a wealth of information that can be used to help determine validity of SFST
- Use Cannabis as example
HGN Clues

<table>
<thead>
<tr>
<th></th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSP</td>
<td>18</td>
</tr>
<tr>
<td>Max Dev</td>
<td>14</td>
</tr>
<tr>
<td>Prior 45</td>
<td>8</td>
</tr>
</tbody>
</table>

<5 ng/ml

5+ ng/ml
One Leg Stand Clues

Percent

- Sway
- Arms
- Hops
- Foot Down

Green bar: <5 ng/ml
Red bar: 5+ ng/ml
Predictive/Criterion Validity

- Use data in manner similar to original SFST validation studies for alcohol

<table>
<thead>
<tr>
<th>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

<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;2 Clues (Not Impaired)</td>
<td>374 (True Negatives)</td>
<td>431 (False Negatives)</td>
</tr>
<tr>
<td>2+ Clues (Impaired)</td>
<td>54 (False Positives)</td>
<td>31 (True Positives)</td>
</tr>
</tbody>
</table>

- **Accuracy** = 46%
- **Sensitivity** = 7% (ability to detect true positive cases)
- **Specificity** = 87% (ability to detect true negative cases)
Walk and Turn

<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>

- Accuracy = 51%
- Sensitivity = 77%
- Specificity = 22%
One Leg Stand

<table>
<thead>
<tr>
<th>OLS 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>221 (True Negatives)</td>
<td>226 (False Negatives)</td>
</tr>
<tr>
<td>2+ Clues (Impaired)</td>
<td>202 (False Positives)</td>
<td>202 (True Positives)</td>
</tr>
</tbody>
</table>

- Accuracy = 48%
- Sensitivity = 47%
- Specificity = 52%
Validity of SFST for Cannabis

- SFST unable to distinguish between THC level 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

<table>
<thead>
<tr>
<th></th>
<th>THC = 0</th>
<th>THC = 1+</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSP</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>Max Dev</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Prior 45</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

Percent
One Leg Stand Clues

Percent

Sway
Arms
Hops
Foot Down

THC = 0
THC = 1+

Page 37
Walk and Turn Clues

- THC = 0
- THC = 1+

<table>
<thead>
<tr>
<th></th>
<th>THC = 0</th>
<th>THC = 1+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Too Soon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lose Bal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heel /Toe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Off Line</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turn</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Percent
### HGN

<table>
<thead>
<tr>
<th>HGN Score</th>
<th>THC = 0</th>
<th>THC 1+ ng/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;2 Clues (Not Impaired)</td>
<td>407 (True Negatives)</td>
<td>1403 (False Negatives)</td>
</tr>
<tr>
<td>2+ Clues (Impaired)</td>
<td>28 (False Positives)</td>
<td>160 (True Positives)</td>
</tr>
</tbody>
</table>

- **Accuracy** = 28%
- **Sensitivity** = 10% (ability to detect true positive cases)
- **Specificity** = 94% (ability to detect true negative cases)
Walk and Turn

<table>
<thead>
<tr>
<th>WAT Score</th>
<th>THC = 0</th>
<th>THC 1+ ng/ml</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>

- Accuracy = 78%
- Sensitivity = 82%
- Specificity = 65%
# One Leg Stand

<table>
<thead>
<tr>
<th>OLS Score</th>
<th>THC = 0</th>
<th>THC 1+ ng/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1 Clues (Not Impaired)</td>
<td>261 (True Negatives)</td>
<td>353 (False Negatives)</td>
</tr>
<tr>
<td>1+ Clues (Impaired)</td>
<td>165 (False Positives)</td>
<td>1204 (True Positives)</td>
</tr>
</tbody>
</table>

- **Accuracy** = 74%
- **Sensitivity** = 77%
- **Specificity** = 61%
<table>
<thead>
<tr>
<th>THC Threshold</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ng/mL</td>
<td>80%</td>
<td>70%</td>
<td>77%</td>
</tr>
<tr>
<td>2 ng/mL</td>
<td>72%</td>
<td>75%</td>
<td>73%</td>
</tr>
<tr>
<td>3 ng/mL</td>
<td>60%</td>
<td>78%</td>
<td>66%</td>
</tr>
<tr>
<td>5 ng/mL</td>
<td>40%</td>
<td>84%</td>
<td>55%</td>
</tr>
<tr>
<td>7 ng/mL</td>
<td>30%</td>
<td>89%</td>
<td>50%</td>
</tr>
<tr>
<td>10 ng/mL</td>
<td>14%</td>
<td>95%</td>
<td>41%</td>
</tr>
</tbody>
</table>
Reliability

Reliability is concerned with consistency in measurement.

Many types of Reliability

1. Inter-rater reliability
2. Test-retest reliability
Reliability

Data Collection
- DRE certification sessions
- 2 evaluators scoring tests 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>
</tbody>
</table>

- **2+ Clues**: 87%
### Reliability

**One Leg Stand**

<table>
<thead>
<tr>
<th>Clue</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

- # Clues: 78%
- 2+ Clues: 87%
Reliability
Horizontal Gaze Nystagmus

<table>
<thead>
<tr>
<th>Clue</th>
<th>Agreement</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSP</td>
<td>96%</td>
<td>(33)</td>
</tr>
<tr>
<td>Max Dev</td>
<td>99%</td>
<td>(21)</td>
</tr>
<tr>
<td>Onset &lt;45</td>
<td>99%</td>
<td>(19)</td>
</tr>
</tbody>
</table>

(most subjects did not show HGN)
SFST & Drugs

- Reliability
  - Inter-rater agreement

- Validity
  - WAT & OLS
  - HGN – limited
  - Can we do better?
Additional Tests/Indicators

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

<table>
<thead>
<tr>
<th>Measure</th>
<th>THC = 0</th>
<th>THC = 1+</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>Romb 2+</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>Eye Trem</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>Droopy</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>FTN 2+</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
### Other Indicators for Cannabis

<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 2+</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>
Reliability
Other Measures

Agreement

- LOC 90%
- Leg Tremors 81%
- Eyelid Tremors 73%
- Romberg F/B 47%
- Romberg S/S 56%
- Romberg (2+) 81%
- Clock (+/- 2) 84%
- FTN Misses 36%
Issues with Other Measures

- Recording
- Scoring/Clues
- Norms
- Impairment
- Standardization
- Validation
- Legal Challenges
Finger to Nose
Alphabet Backwards

- Scoring
- Norms
- Validation
- Learning Effect
International Practices

- **Victoria, Australia**
  - SFST part of Preliminary Impairment Test

- **United Kingdom**
  - Field Impairment Test (FIT)
  - Includes WAT, OLS, Romberg, FTN, Pupil size

- **New Zealand**
  - Compulsory Impairment Test (CIT)
  - HGN, WAT, Romberg, Pupil Size, OLS, FTN
International Practices

- Germany
  - Checklist of signs and symptoms
  - Based in part of DEC program

- Norway
  - Simplified Clinical Test for Impairment (SCTI )
  - Includes WAT, FTN, Romb, FTF
  - (HGN reserved for assessment by physician)
In conclusion…

- There is evidence to support the reliability and validity of SFST to detect drug use in drivers
- Only looked at cannabis…
- HGN specific to drug category
- May want to consider additional tests/indicators
- Work is ongoing…
Doug Beirness
DBeirness@magma.ca
613-820-5236