Screening Tests for Driving with a Focus on cognition

Michel Bédard, PhD
Director, Centre for Research on Safe Driving
Professor, Lakehead University & Northern Ontario School of Medicine
Scientific Director, Centre for Applied Health Research - St. Joseph’s Care Group
Deaths among motor vehicle occupants

Deaths per 100,000 population

Ramage-Morin, 2008
Frailty bias*

*Controlled for sex, BAC, site of impact, restraint use, traveling speed, vehicle model year, vehicle wheelbase (Bédard et al., 2002)
Canada

Fatalities

Year

Total fatalities

1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

Transport Canada, 1999-2010
Fatality trends: USA

Mullen, Dubois & Bédard, 2013
At-Fault Collision Involvement Rates By Driver’s Age, 2002-2010*

*Courtesy of MTO; Includes drivers of light duty vehicles only, involved in fatal, personal injury and property damage collisions
Low mileage bias (Langford et al., 2006)

Fig. 2. Annual crash involvement for different driver ages, controlling for annual mileages.
Age matters!

Bédard, 2000
The driving task (Michon, 1979, 1985)

• Strategic level
  • Advanced decision-making (e.g., choosing a route, staying home in bad weather)
  • Requires a mix of judgement and memory

• Tactical level
  • Decision-making while driving (e.g., speed, distance between vehicles)
  • Requires scanning of environment, awareness, and anticipation

• Operational level
  • Actual maneuvers (e.g., turning, braking)
  • Requires perception, ability to act adequately
Properties of a good screening test

- Inexpensive
- Easy to administer
- Safe
- Minimal discomfort/harm
- High level of acceptability
- High level of reproducibility and validity – the test consistently and correctly identifying individuals with the condition and those without
Validity

- Usually based on a “gold standard”, accepted criterion to determine presence of the condition (e.g., on-road test, crashes)
- Also “face validity”, appears to measure what it is supposed to
- How do we measure if it correctly classifies?
Validity

- Examine simple measures of association (e.g., correlation, relative risk, odds ratio) between a predictor and the gold standard
  - Is there an association beyond the play of chance?
- Many tests are statistically associated with driving outcomes. Some examples:
  - Mini-mental State Examination (MMSE)
  - Trail making test (versions A & B)
  - MVPT (motor-free visual perception test)
  - UFOV©
Important considerations

• $P$-values are heavily influenced by sample sizes
  • Small associations may be beyond the play of chance yet have little clinical/practical importance

• The magnitude of the association (e.g., $R^2$, , RR, OR) tells an important part of the story

• Associations at the group level, unless they are perfect, do not necessarily translate into good prediction at the individual level
Some claims

• “the UFOV test is a valid indicator of retrospective and concurrent driving performance in older adults [and the] . . . evidence of a relationship between UFOV test performance and driving performance has far-reaching implications for public policy. (Clay et al., 2005, p. 73)

• The SIMARD-MD is a tool “with a high degree of accuracy that can be used for immediate decisions in the clinical setting.” (Dobbs and Schopflocher, 2010, p. 119)
More claims

• “DriveABLE provides a science based cognitive assessment solution to identify those drivers who have become unsafe to drive due to medical impairments.”
  [Link](http://driveable.ca/resources/Science_Behind_DriveABLE.pdf)

• “The findings provide the evidence physicians need to be confident in using the recommendations from the DriveABLE In-Office cognitive evaluation...” (p. e161; Dobbs, 2013)
Different worlds

- **Theoreticians**: academic researchers focused on development & testing of theories about how things work
  - Typically report *group level* effects

- **Practitioners**: here, referring to clinicians making decisions about individuals’ fitness to drive (e.g., physicians, occupational therapists, optometrists)
  - Clinicians need fairly strong evidence that applies to *individual* patients or clients
## The basic situation

### Test Result

<table>
<thead>
<tr>
<th></th>
<th>Unfit</th>
<th>Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive</strong></td>
<td>![Thumbs Up]</td>
<td>![Thumbs Down]</td>
</tr>
<tr>
<td>(unfit to drive)</td>
<td>$a = TPF$</td>
<td>$b = FPF$</td>
</tr>
<tr>
<td><strong>Negative</strong></td>
<td>![Thumbs Down]</td>
<td>![Thumbs Up]</td>
</tr>
<tr>
<td>(fit to drive)</td>
<td>$c = FNF$</td>
<td>$d = TNF$</td>
</tr>
</tbody>
</table>

### Actual Fitness-to-Drive

- **Positive Predictive Value (PV+):** $PV+ = \frac{a}{a + b}$
- **Negative Predictive Value (PV−):** $PV− = \frac{d}{c + d}$

### Sensitivity

- $= \frac{a}{a + c}$

### Specificity

- $= \frac{d}{b + d}$
Sensitivity

Sensitivity* = the proportion who truly are UNFIT to drive that are correctly labeled as UNFIT by the test.

* Calculated as “a/(a+c)”; also known as the True Positive Fraction (TPF)
Specificity* = the proportion who truly are *FIT to drive* that are correctly labeled as *FIT by the test*.

* Calculated as “d/(b+d)”; also known as the True Negative Fraction (TNF)
AUC = area under the curve

ROC Space

Perfect discrimination; AUC = 1.0

Better

No discrimination, or random guess; AUC = 0.5

Worse

AUC = area under the curve

TPF or sensitivity

FPF or (1 – specificity)
TPF OR sensitivity

FPF or (1 - specificity)

- UFOV Divided Attention
- SMMSE
- Trails A
- # of prior driving offenses
Using two cut-off values

• To obtain adequate sensitivity and specificity it may be necessary to use two cut-off values.

Test Result
Clearly Positive       Grey Zone       Clearly Negative
--------------------------------------------------------------------
A1        A2        A3  ......................  B3       B2       B1
Must trade off between the need for adequate sensitivity and specificity, the size of the grey area, and the cost of making false determinations.
A1: below this value all drivers are unfit (based on fit data)
B1: above this value all drivers are fit (based on unfit data)

Using A1 and B1 you have a large grey area

A2: below this value some fit drivers will be classified as unfit (false +)
B2: above this value some unfit drivers will be classified as fit (false -)
The SIMARD-MD approach*

- SIMARD ≤ 30 (A2)
  - Stop

- SIMARD 31-70
  - DriveAble*

- SIMARD > 70 (B2)
  - Drive

*Based on Dobbs & Schopflocher, 2010
**DriveAble is a for-profit company; decision is also based on a trichotomy following the in-office test
Their data

<table>
<thead>
<tr>
<th>Cut-off</th>
<th>Failed on-road</th>
<th>Passed on-road</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤30 (A2; 27%)</td>
<td>80%</td>
<td>20%</td>
</tr>
<tr>
<td>31-70 (50%)</td>
<td>52%</td>
<td>48%</td>
</tr>
<tr>
<td>&gt;70 (B2; 23%)</td>
<td>13%</td>
<td>87%</td>
</tr>
</tbody>
</table>
SIMARD-MD

Bédard et al., 2011
Issues to consider

- Clinicians/assessors should:
  - Insist on measures of prediction in addition to measures of association
  - Insist on seeing the actual data and ROC curves to identify potential cut-off points and the size of grey areas
  - Not take p-values and statistical associations as a sign that a test can predict safe driving
  - Focus on 100% sensitivity and specificity cut-off values
- Framework published by Dickerson and Bédard (2014)
- Be aware of education biases
Need to adjust for education effect*

<table>
<thead>
<tr>
<th>Test (n = 244)</th>
<th>% of mean</th>
<th>B (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trails B</td>
<td>14.76</td>
<td>-14.42 (-20.57, -8.27)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SIMARD</td>
<td>11.47</td>
<td>8.19 (4.99, 11.40)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Trails A</td>
<td>7.35</td>
<td>-2.91 (-4.95, -0.87)</td>
<td>.005</td>
</tr>
<tr>
<td>MVPT</td>
<td>6.06</td>
<td>-7.46 (-13.70, -1.22)</td>
<td>.019</td>
</tr>
<tr>
<td>MoCA</td>
<td>2.87</td>
<td>0.74 (0.37, 1.12)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CRSD-ANT</td>
<td>2.43</td>
<td>-19.19 (-38.61, 0.22)</td>
<td>.053</td>
</tr>
<tr>
<td>MMSE</td>
<td>1.90</td>
<td>0.54 (0.32, 0.76)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*Based on linear regression models including age (centered at 75 with 5-year increments) and education (no post-secondary as referent). Bédard et al., 2013
Driving is an ADL

- Not different than many others
- Hierarchy: flying > driving > preparing a meal
- Many tools used to assess other ADLs and cognition can provide insight into fitness-to-drive
Issues to consider

- No single tool should be used as the basis to make a decision about the driving privilege (see Dickerson et al., 2014 for a review)

- The evidence behind tools is not always solid and publication in a journal (or even adoption by a jurisdiction) is not sufficient to take the authors recommendation at face value (see Weaver & Bédard [2012] “Tips on choosing fitness-to-drive screening tools”)

- Other approaches (e.g., serial trichotomization, risk stratification tool [Candrive], probability calculator [Barco et al.])
Back to the drawing board

- Balance between public safety and privilege to drive
- Decision to revoke a license based on valid and reliable measures and process
- Gold standard to assess fitness to drive = On-road evaluation by clinician with specialized training
- But!
  - Off-road tests more efficient and safer
Limitations of off-road tests

- Individual tests not precise in clinical setting
- Acceptable level of Sensitivity and Specificity?

Molnar and colleagues (2009) proposed:

Serial Trichotomization
Many clinicians may prefer to start with cognitive tests. When physicians employ cognitive tests such as the MMSE, Clock Drawing and/or Trails A and B, they should keep in mind that none of these have well-validated cut-off scores. In the case of overlapping or unclear cognitive scores, serial triaging of test results (e.g., clearly unsafe, uncertain with further testing required, no concerns regarding safety) can facilitate judgment of driving fitness. The unclear category may be further evaluated by considering qualitative dynamic information regarding how the test was performed (e.g., observations such as slowness, hesitation, multiple corrections, anxiety, impulsive or perseverative behavior, lack of focus, forgetting instructions, inability to understand test, etc.), which may help in the interpretation of this category of patients. The triaging approach essentially asks: "Which patients are obviously unfit to drive, which are clearly safe, and which require further evaluation?"
Our data

• Obtained by occupational therapist (OT) with specialized training in driver assessment

• Cognitive measures:
  a. Clock Drawing Test (CDT)
  b. Montreal Cognitive Assessment (MoCA)
  d. Trail Making Test Part A (Trails A)
  e. Trail Making Test Part B (Trails B)

• Followed by on-road driving test
Statistical Analysis

1. Calculated receiver operating characteristic (ROC) curves and area under the curve (AUC) for each test
2. Used the ROC curve to identify upper and lower cut-points that achieved 100% sensitivity and 100% specificity in predicting the on-road driving test outcome
3. Participants’ test scores categorized into:
   - Pass (test scores >100% sensitivity cut-point)
   - Fail (test scores <100% specificity cut-point)
   - Indeterminate (test scores between the cut-points)

4. Serial trichotomization used to filter participants into Pass or Fail categories
## Results

<table>
<thead>
<tr>
<th>Test</th>
<th>Cut-point for 100% Sensitivity</th>
<th>Cut-point for 100% Specificity</th>
<th>% Indeterminate</th>
<th>Area Under the Curve (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trails B (seconds)</td>
<td>80</td>
<td>178</td>
<td>33.7</td>
<td>0.979 (.948, 1.000)</td>
</tr>
<tr>
<td>CDT</td>
<td>7</td>
<td>4</td>
<td>63.4</td>
<td>0.841 (.744, .937)</td>
</tr>
<tr>
<td>MVPT-3</td>
<td>57 (37)*</td>
<td>41 (26)*</td>
<td>65.1</td>
<td>0.922 (.861, .983)</td>
</tr>
<tr>
<td>Trails A (seconds)</td>
<td>25</td>
<td>69</td>
<td>66.3</td>
<td>0.914 (.848, .981)</td>
</tr>
<tr>
<td>MoCA</td>
<td>27</td>
<td>16</td>
<td>80.7</td>
<td>0.889 (.814, .964)</td>
</tr>
</tbody>
</table>

* The MVPT-3 values in brackets represent the pro-rated values assuming a 42-point scale.
Serial trichotomization
Implementation

• Complete tests in sequence until there is a determination of “pass” or “fail”.
  • If client fails Trails B, allow to continue but a second fail means he/she is determined unfit to drive
• Adjusted cut-off values to account for education bias on Trails A and B
  • Two different sets of cut-off values: 1) clients with high school education or less, and 2) clients with post-secondary education
Limitations

- Not definitive
- Small sample
- Spectrum bias
- Incorporation bias
- Need for more research (prospective validation)
- Beware of situations where the driver may be able to perform better than he/she did (e.g., tool bias, reversible situation)
- Referral to specialists is crucial to deal with borderline situations or drivers who may have potential for remediation
Thank You!

Supported by

- Huge number of colleagues
- Canada Research Chair Program, Government of Canada
- Networks of Centres of Excellence (AUTO21)
- Canadian Institutes of Health Research (CIHR)
- Ontario Neurotrauma Foundation (ONF)
- Natural Sciences and Engineering Research Council (NSERC)
- Ontario Ministry of Energy, Science and Technology
- Canada Foundation for Innovation (CFI)
- Thunder Bay Community Foundation
- Lakehead University
- St. Joseph’s Care Group

http://crsd.lakeheadu.ca/
Some resources


Predictive value of a positive test (PV+)

PV+* = the proportion who truly are \textit{UNFIT to drive} among all those labeled as \textit{UNFIT by the test}.

* Calculated as “a/(a+b)”
Predictive value of a negative test (PV-)

PV-* = the proportion who truly are **FIT to drive** among all those **labeled as FIT by the test**.

* Calculated as “d/(c+d)”
<table>
<thead>
<tr>
<th></th>
<th>%</th>
<th>Fail</th>
<th>Border.</th>
<th>Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfit (A2)</td>
<td>74.5</td>
<td></td>
<td>11.7</td>
<td>13.9</td>
</tr>
<tr>
<td>In-office</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind.</td>
<td>40.1</td>
<td></td>
<td>24.0</td>
<td>35.9</td>
</tr>
<tr>
<td>Fit (B2)</td>
<td>12.3</td>
<td></td>
<td>19.6</td>
<td>68.1</td>
</tr>
</tbody>
</table>
### Demographic characteristics

Participants (n=83)

<table>
<thead>
<tr>
<th>Age – Mean (SD)</th>
<th>60.78 (15.36)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex – Female, n (%)</td>
<td>21 (25.3)</td>
</tr>
<tr>
<td>Reason for Referral – n (%)</td>
<td></td>
</tr>
<tr>
<td>• cerebrovascular accident</td>
<td>32 (38.6)</td>
</tr>
<tr>
<td>• traumatic brain injury</td>
<td>13 (15.7)</td>
</tr>
<tr>
<td>• cognitive deficits</td>
<td>10 (12.0)</td>
</tr>
<tr>
<td>• amputation of limb</td>
<td>6 (7.2)</td>
</tr>
<tr>
<td>• spinal injury</td>
<td>3 (3.6)</td>
</tr>
<tr>
<td>• multiple sclerosis</td>
<td>3 (3.6)</td>
</tr>
<tr>
<td>• other (e.g., mental illness, chronic pain)</td>
<td>16 (19.3)</td>
</tr>
</tbody>
</table>
## On-road test results

### Outcome of On-Road Test – n (%)

<table>
<thead>
<tr>
<th>Fit to Drive</th>
<th>55 (66.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfit to Drive</td>
<td>28 (33.7)</td>
</tr>
</tbody>
</table>
## Descriptive statistics

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trails B (seconds)</td>
<td>158.14</td>
<td>139.50</td>
</tr>
<tr>
<td>CDT</td>
<td>5.61</td>
<td>1.46</td>
</tr>
<tr>
<td>MVPT-3</td>
<td>51.07</td>
<td>7.69</td>
</tr>
<tr>
<td>Trails A (seconds)</td>
<td>46.94</td>
<td>33.20</td>
</tr>
<tr>
<td>MoCA</td>
<td>23.63</td>
<td>4.28</td>
</tr>
</tbody>
</table>