CAN OFF-ROAD SCREENING TESTS ASSESS OLDER DRIVERS' FITNESS TO DRIVE?

Jim Langford1,2, Jude Charlton2 and Megan Bohensky2

1 Curtin-Monash Accident Research entre, Western Australia
2 Monash University Accident Research Centre, Clayton Victoria

Abstract

The Austroads older driver licensing model calls for a quick and convenient, valid screening test to provide an early assessment of fitness to drive. This paper presents the evidence from two key sources regarding the usefulness of screening tests in this context. First, the research literature has been examined to assess the validity and usefulness of some leading screening tests being used overseas. Secondly, results from a recently-completed case-control study evaluating the capacity of screening measures to distinguish between safe and unsafe drivers are presented. Based on these results, it has been concluded that off-road screening tests of fitness to drive currently cannot be justified for mass, age-based screening for licensing decisions.

Keywords

Older driver safety; screening tests; assessment of fitness to drive.

1. INTRODUCTION

1.1 The calls for a valid screening test to assess fitness to drive

It is generally accepted that some older drivers represent an unacceptable crash risk. Arguably, assessment procedures should focus only upon those drivers giving some preliminary evidence of being at risk, as reflected in the Austroads licensing model (Fildes et. al, 2000). The model’s features include:

- the establishment of a network of community notification sources, whereby only drivers suspected to have a high crash risk are referred to the licensing authority for formal assessment;
- the use of multi-tiered assessment, involving off-road screening tests of fitness to drive and then if appropriate, more elaborate levels assessment procedures;
- an emphasis on assessing functional abilities for safe driving; and
- the use of assessment instruments of acceptable validity for assessing crash risk.

The call for valid off-road tests of functional capacity has also been made overseas, albeit within a different licensing context to that proposed in the Austroads model. A recent report to the National Highway Traffic Safety Administration in the United States recommended that all drivers reaching a designated age – notionally 70 to 75 years – be assessed every two years, with assessment to include a functional abilities screening test (Lococo and Staplin 2005). This call was repeated by an expert group at the North American License Policies Workshop (AAA Workshop, 2008):

Although it is not appropriate (or practical) to have age-triggered assessment, it is appropriate to have age-triggered driver screening, with screening only used to see if further testing should be done ….

However the Workshop participants did qualify this recommendation:

Screening and assessment tools used in licensing settings must be valid and reliable, and also efficient, easily adopted, and cost effective. They need to balance scientifically sound means to identify potentially at-risk drivers against the practical limitations and cost of what is reasonable for a licensing agency to accomplish, while at the same time treating individuals in a fair and dignified manner.

1The distinction between screening and assessment is critical. As reflected in the AAA Workshop citation: screening is intended to provide only a preliminary measure of a given quality and in some instances is likely to lead to more rigorous testing; assessment entails more rigorous testing and leads to categorical decisions being made in regard to the given quality.
The extent to which current screening tests meet these criteria remains open to discussion.

1.2 How valid are the current screening tests?

Arguably the best credentialed screening test is the Useful Field of View (UFOV), first described by Ball et al. (1993). Seven major studies have since been conducted by the test developers, assessing aspects of the association between UFOV performance and crash involvement (Ball et al. 1993, Owsley 1994, Goode et al. 1998, Owsley et al. 1998b, Owsley et al. 1998a, Sims et al. 1998, Sims et al. 2000). UFOV has also been evaluated by many other researchers independent of the initial research team. Rubin et al. (2006) as a recent example, conducted a series of vision tests (including UFOV) on a sample of 1801 drivers aged 65-84 years from 1993-95 and also collected crash data over an extended period to 1997. It was found that drivers with a 40% or greater reduction in useful field of view as measured by UFOV, had a 2.12 times risk of crash involvement relative to those with no or acceptable loss of useful field of view.

Performance on UFOV has been consistently and statistically associated with crash involvement, either retrospectively or more impressively, prospectively. This association, while a necessary first step in identifying at-risk drivers, is not enough in itself for a test to be acceptable to and useable by licensing authorities. As a next minimum requirement, any association between test performance and crash risk needs to be quantified.

Empirically- and statistically-sound studies relating UFOV performance to crash involvement have generally reported around a 2-times relative risk of crash involvement (for an overview, see Langford, 2008). It is contended that in applying the tests to all drivers reaching a threshold age, relative risks of this magnitude have little value to licensing authorities and others responsible for the management of older driver safety. The usefulness of UFOV to road safety authorities is further limited when the numbers of ‘safe’ drivers who would be needlessly failed based on ‘unsafe’ UFOV performance, are considered. Langford (2008) has estimated that for every crash saved, 99 drivers who would have driven without crashing in the year following assessment would have lost their licence using conventional UFOV fail criterion in a mass older driver screening context.

Perhaps most tellingly, a US research team which included one of the UFOV developers, concluded that:

“the current screening instruments do not discriminate future crash risk well enough in this primarily low-risk driving population to justify screening all drivers aged 75 years and older. … a near-perfect screening test would be required in order for screening to be a worthwhile option.” (Viamonte et al. 2006, pp. 356-7).

At least in some quarters, UFOV is now being treated primarily as a training rather than assessment procedure (AAA, 2009).

1.3 Purpose of this paper

Adding to the findings from the research literature, this paper presents further evidence from a study conducted in Western Australia regarding the usefulness of screening tests in identifying at-risk older drivers.

2 METHOD

2.1 Aim of the study

The overall aim of the study was to evaluate the capacity of each of sixteen assessment protocols to distinguish between older drivers recently involved in at-fault crashes and older drivers not involved in crashes.
2.2 Summary of research design

A case-control study design was employed, comprising 62 Cases and 62 Controls. Each Case was aged at least 75 years, resident in or near to the Perth metropolitan area, involved in a recent casualty crash and judged by insurance assessors to have been responsible for the crash. Controls were matched as closely as possible for age, gender and area of residence and had a crash-free history.

The sixteen assessment protocols were:

- **Health Screen for Drivers (HSD) (Charlton et al., 2003):**
  1. total score; and component item scores;
  2. Visualisation of Missing Information (VMI) Test;
  3. Months Backwards;
  4. Confusion rating;
  5. Slowness rating;

- **Autotrails:**
  6. Ascending;
  7. Descending;

- **UFOV®**
  8. overall UFOV®
  9. subtest 1;
  10. subtest 2;
  11. subtest 3;

- **Driver self ratings:**
  12. driving ability compared to most other drivers;
  13. driving ability compared to drivers aged 30 to 50 years;
  14. driving ability compared to 20 yrs ago;
  15. occupational therapist rating;

- **Driving performance**
  16. On-road rating by a driver assessor.

The primary analyses reported in this paper were based on the association between prior crash involvement and the determination of whether the assessment was passed (Y/N). In addition, a tri-category assessment protocol was used for selected HSD and UFOV scores, so that these tests would set two thresholds to distinguish three categories. These analyses were restricted to nine assessment protocols which could most readily serve as screening tests. This approach was adopted for consistency with the approach proposed for use of screening tests in the Australasian older driver re-licensing model (Fildes et al., 2004) and as a possible means to improve test sensitivity and specificity (Molnar et al., 2006). Ideally, a screening test should be able to predict safe and unsafe drivers with 100% certainty. In practice, however, this is rarely achieved given test and individual variation. In the tri-level assessment model, any result below the lower score would be a fail, any result above the upper score would be a pass and those in-between would require further testing, involving health care providers and/or on-road tests. Threshold scores were identified to optimise the tests’ ability to identify correctly both unsafe and safe drivers while maintaining manageable numbers of doubtful cases requiring further assessment.

Details of each protocol, the scoring procedures and the various Pass, Fail and Doubtful thresholds are given in Langford et al. (in press).

3 RESULTS

3.1 Test performance on a simple pass/fail basis

Cases and Controls differed in performance on twelve of the sixteen assessment protocols and the difference was in the expected direction: more Cases than Controls failed. These protocols were:

- Months Backwards;
- Confusion rating;
- Slowness rating;
For the remaining four protocols, the results were in an unexpected direction: more Controls than Cases failed. These protocols were:

- VMI Test;
- Autotrails Descending;
- UFOV® subtest 1; and
- self-rated driving ability compared to most other drivers.

However these results do not endorse the validity of the screening protocols. Based on a two-level (pass/fail) classification of screening test performance outcomes, Cases and Controls did not differ significantly on any of the measures, with most of the differences in the expected directions amounting to only a very small number of drivers.

3.2 Test performance to produce three-level outcomes

Results from nine test items were further analysed to allocate drivers to one of three categories; ‘pass’, ‘maybe’ and ‘fail.’

As was the case for simple Pass-Fail analyses, valid screening tests need to meet two basic criteria: a greater proportion of Cases (crash involved) need to be identified as Fails, and a greater proportion of Controls (not crash involved) need to be identified as Passes - with both Cases and Controls who could not be confidently identified as either Passes or Fails being categorised as Maybe. In summary:

- four tests (Months Backwards, Ascending Autotrails, UFOV® sub-test 2 and the Combined UFOV®) met both criteria;
- two tests (HSD and UFOV® sub-test 3) met one criterion only; and
- three tests (Motor-Free Visual Perception Test, Autotrails Descending and UFOV® sub-test 1) failed to meet both criteria.

With one exception (Ascending Autotrails), none on the differences between Cases and Controls was statistically significant.

4 DISCUSSION - VALIDITY OF THE SCREENING ASSESSMENT PROTOCOLS

Based on Cases’ and Controls’ performances, the results obtained in the Western Australian study do not support any of the sixteen screening protocols as being useable for mass, age-triggered assessment purposes.

There are several possible explanations for the absence of significant differences between Cases and Controls. In particular, the study relied upon voluntary participation of both Cases and Controls, with the 62 Cases representing 24% of all Cases invited to participate. It may well have been that a self-selection bias was operating whereby only those Cases who were relatively confident of their fitness to drive

---

2 In practice and to be useful to licensing authorities, screening tests need to be comply with these two criteria AND also have adequate sensitivity and specificity to ensure that adequate proportions of unsafe and safe drivers are correctly categorised.
volunteered. If their confidence were justified, the result would be Cases who represented the fittest of older drivers in recent at-fault crashes. This possible bias would reduce the magnitude of any assumed differences between Cases and Controls.

There are other possible explanations. It may also be that Cases’ recent involvement in at-fault crashes:

- was due to different exposure factors. Based on self-reported driving activity, while both groups drove almost exclusively in urban areas, Cases tended to drive more frequently and greater distances than Controls, although the differences were not significant. If Cases’ crashes had resulted even partly from different driving patterns and higher exposure levels rather than from functional deficits, there would be less chance of obtaining differences in performance on the assessment protocols;

- had no association with differences in functional decline but was due to different compensatory factors. Hennessy and Janke (2005) have developed the notions of ‘driving wellness’ which relates to individual factors relevant to safe driving (including functional health) and ‘driving fitness’ which relates to how well the individual actually drives (and may include strategies to compensate for any reduced functional health). Perhaps Cases and Controls both had comparable ‘driver wellness’ but Cases compensated less for possible limitations;

- was due to functional deterioration in critical areas not tapped by the sixteen measures used in this study; Functional decline in older drivers may be also be due to a variety of different factors and conditions and one “catch-all” assessment protocol is not going to adequately detect everything.

- was due to functional decline which may be sporadic, evident at the time of crash but not evident at the time of assessment. It is possible that older drivers, especially but not only with cognitive decline, might compensate adequately under most circumstances but might make critical errors in high-pressure situations. Perhaps many Cases’ pre-crash scenarios represented high-pressure situations, whereas the assessment situations did not. If this possibility were true, again, Cases and Controls would not be expected to differ on the assessment protocols;

- was due to chance. Perhaps the only substantial difference between Cases and Controls was that the former were simply unlucky. Their errors resulted in crashes, whereas Controls making perhaps the same errors escaped this outcome.

The discussion to this point has centred upon assessment protocols and the statistically significant association (or the lack thereof) between driver safety status and performance. This association is a necessary first step if a test is to have value in identifying at-risk drivers. However demonstrating this association is not enough in itself for a test to be acceptable to and useable by licensing authorities. As a next minimum requirement, an assessment protocol needs to distinguish between safe and unsafe drivers with sufficient sensitivity and specificity to satisfy all key parties. However the failure of the sixteen assessment protocols used in this study to achieve statistical significance as a first step, makes further discussion of required sensitivity and specificity for the purposes of this paper unnecessary.

5. CONCLUSIONS AND RECOMMENDATIONS

No evidence was found either from the research literature or from the Western Australian study that existing screening protocols are able to make a meaningful distinction between safe and unsafe drivers, and therefore it is unlikely that they will be useful to licensing authorities. More specifically, it follows that in the absence of more valid and reliable assessment protocols, the Austroads older driver licensing model is currently inoperable. However in reaching this conclusion, it also needs to be recognised that the mandatory licensing assessments currently being conducted around the world and in eight of the nine Australasian jurisdictions similarly lack demonstrable validity – and may even be decreasing the overall quality of older drivers on the road. That is some proportion of Fails will be incorrect, resulting in unfair licence withdrawals and further, a proportion of safe drivers will (inappropriately) surrender their licence rather than undergo the stressful process of re-testing (Oxley et al., 2003).

Given the Austroads model’s potential to improve older driver safety, it is recommended that the development of more effective assessment protocols and associated processes be continued. The proposed MUARC-led involvement in an international older driver cohort study, which is aiming to
identify/develop and evaluate a battery of both on-road and off-road assessment protocols, is therefore welcomed.

6. REFERENCES


Langford J, (2008), Usefulness of off-road screening tests to licensing authorities when assessing older driver fitness to drive, Traffic Injury Prevention. 9, 328-335.


