TRL

PUBLISHED PROJECT REPORT PPR061

THE VALIDITY AND RELIABILITY OF THE STROKE DRIVERS SCREENING ASSESSMENT

Version: 1

by J Sentinella (TRL) and L Read (University of Leeds)

Prepared for: Project Record:	**PPAD 9/72/62 Validity and Reliability of the Stroke Drivers' Screening Assessment**
Client:	**Mobility and Inclusion Unit, Department for Transport (Yvonne Brown)**

	Approvals
Project Manager	
Quality Reviewed	

This report has been produced by TRL Limited, under/as part of a Contract placed by Department for Transport. Any views expressed are not necessarily those of Department for Transport.

TRL is committed to optimising energy efficiency, reducing waste and promoting recycling and re-use. In support of these environmental goals, this report has been printed on recycled paper, comprising 100% post-consumer waste, manufactured using a TCF (totally chlorine free) process.

CONTENTS

Executive summary

A stroke can cause physical and cognitive impairments such as lack of concentration, visual inattention and reasoning difficulties which may affect a drivers' ability to drive safely. The Stroke Drivers Screening Assessment (SDSA) has been developed by the University of Nottingham as a screening tool to identify drivers who are not fit to resume driving following a stroke. TRL and the University of Leeds were commissioned by the Department for Transport to investigate the current use of the SDSA in the UK and examine the validity and reliability of the screening tool.

The project consisted of four studies:

- A survey of SDSA users establish the extent to which the SDSA is currently being used across the United Kingdom; consider the ease of its use; and investigate the advice given based on the results of the test.
- An inter-rater reliability study to investigate whether different assessors reach the same outcome and whether the scoring procedure and the interpretation of the results are accurate and rigorous.
- A test-retest reliability study to investigate whether a person tested on two separate occasions will obtain the same result and whether any possible learning effects resulting from repeated administration of the test are likely to lead to a change in the outcome (from fail to pass).
- A validity study to investigate the criterion-related validity (predictive validity and concurrent validity) and face validity of the SDSA. The criterion used was performance in an on-road driving assessment. To gain an indication of concurrent validity the Visual Attention Analyser Useful Field of View (UFOV) test was used as measure.

The survey of users suggests that the SDSA is not widely used and those healthcare professionals who are currently using the SDSA do so infrequently and not routinely with all patients. There appears to be little or no consensus among users on the circumstances under which the SDSA should be used and this is an issue that needs to be addressed.

Most of the current users who responded to the survey were occupational therapists, who felt the SDSA is 'useful' in helping to assess a client's fitness to drive but that on its own cannot be used to predict driving performance. It is, however, regarded by users as a good indicator of those clients that should definitely not persist in returning to drive and those who need to undergo a full driving assessment. The raters who participated in the inter-rater reliability study said they would not use the outcome of the SDSA alone to advise clients that they should not return to

driving. They would instead recommend further assessment, whether that be physical, cognitive or a driving assessment.

The SDSA was believed to be easy to administer and score, although some changes to the SDSA manual, the provision of an overlay for the dot cancellation test and an electronic version of the score sheet which automatically calculates the pass and fail equations could improve accuracy.

The SDSA had high inter-rater reliability for each of its components. The interpretation of the results, however, was less consistent than scoring and was found to vary between raters in the inter-rater reliability study. A mixed outcome was found for two clients whose overall SDSA outcome score was close to the pass-fail threshold, with some raters either passing or failing the client and others using a 'borderline' category. However, this category was devised by the users themselves as there is no advice in the SDSA on what a borderline category constitutes.

The test-retest study involved a very small sample of participants: only six volunteered for this part of the study. Four of the six participants passed the SDSA on first administration. In one of the two cases where SDSA was failed on first administration, the volunteer passed on second administration. Reasonable levels of agreement between scores on first and second administration for the dot cancellation time, dot cancellation test errors and square matrix test were found. A lower level of agreement was found for the square matrix compass test and road sign recognition test reflecting changes in the rank order of participants at retest. Changes in mean scores, with the scores improving either due to training or familiarity, for each SDSA component were observed for this small sample between first and second administration and second and fifth administration (post-training). However, for all volunteers, the pass-fail outcome of the second administration of the SDSA was identical to the outcome of the fifth administration, despite the fact that feedback on performance had been given for the intervening administrations.

The validity study indicated that the SDSA can identify and screen out a useful proportion of people who would fail the on-road driving assessment. However, if the SDSA were introduced in its present form as a gatekeeper to the on-road assessment, then at least 30% of the people it failed would otherwise have passed the on-road assessment but would have their access to that assessment denied or delayed.

If the pass-fail criterion for the SDSA is adjusted to make it easier to pass, then the study shows that it may be possible to achieve an acceptably low incorrect failure rate while still screening out a useful proportion of people who would go on to fail the on-road driving assessment. However, the limited sample size means that the rate of incorrect failures remains very uncertain and may, therefore, turn out to be unacceptable even with the adjusted pass-fail criterion.

Face validity of the SDSA was also highlighted as an issue in the survey of SDSA users and by participants in the validity study. It appears many clients do not understand how a pen and paper based test can assess their driving ability. It appears the criticism may relate to all tests that do not directly observe driving; not just those used in the SDSA. It is likely that if the SDSA were used as a gatekeeper to the road test, some drivers would question the decision based upon the SDSA, especially if the outcome is negative.

Overall, this study has highlighted a number of concerns about the validity and reliability of the SDSA. Despite being relatively easy to administer and score, the tool is not always administered in a standardised way or interpreted consistently. The validity study has indicated that the predictive validity is inadequate for the tool to be used as a gatekeeper to the on-road driving assessment since in many cases would deny or delay clients' access to a driving assessment that they would otherwise pass. The results show that an adjustment to the pass-fail cut off has the potential to reduce the number of incorrect failures to a more acceptable level while still retaining a useful ability to identify and screen out drivers who would fail the road test. However, a larger dataset would be needed to establish this.

The current performance of the SDSA may be considered sufficient if the implications were less for drivers who fail it. If a driver who failed the SDSA was referred to a driving assessment centre after 3 months rather than being reassessed again using the SDSA, it may be acceptable that 30% or more of the people to fail the SDSA would in fact pass the on-road test. Drivers who fail the SDSA but could pass the on-road assessment would have their access to the on-road assessment delayed rather than denied. If a driver has to pass the SDSA before being referred for an on-road assessment, a driver who is incorrectly failed could be repeatedly denied access to an on-road driving assessment for some months; in this situation the current performance of the SDSA as the gatekeeper to the on-road assessment is probably not acceptable.

These findings are not to say that the SDSA should be abandoned. It may well be that use of the SDSA in combination with other cognitive and physical test may provide additional information to help practitioners decide who to refer for a driving assessment at a Mobility Centre. The survey of SDSA users highlighted that many users already use the tool in this way. The behaviour of the client whilst undertaking the test could also be noted as this may help the practitioner assess the person's abilities.

The revised SDSA manual includes a number of improvements but may benefit from another review to ensure practitioners have sufficient information to administer and interpret the tests in a standardised way. Training could also be provided to users to ensure they understand how to administer the tests.

1 Introduction

A stroke occurs when part of the brain suddenly becomes unable to function because it is starved of oxygen. This happens when the blood supply to the brain is affected because a cerebral artery is blocked either by a blood clot or by a spasm of a blood vessel (cerebral infarction) or if the artery bursts (cerebral haemorrhage).

In England and Wales, around 266,000 people will have a stroke each year (Mant et al., undated). However, around a third of those affected by stroke go on to make a good recovery through rehabilitation exercises with the assistance of occupational and speech therapists, physiotherapists and other professionals.

Strokes can result in both physical and cognitive impairments that could affect a driver's ability to drive safely. These impairments include slowed reactions, perceptual disorders, visual inattention, concentration and reasoning difficulties. The driving task requires the driver to be attentive, to concentrate, to make judgements quickly in rapidly changing circumstances, and to be able to cope when the situation is demanding.

Drivers affected by a stroke are legally required to inform the DVLA of their condition. DVLA's general guidance is that driving must cease for at least a month, but may resume soon after this, if clinical recovery is satisfactory. It is the responsibility of the DVLA Medical Branch to judge a driver's fitness to drive. In making its judgement, the DVLA may seek information from the driver's General Practitioner (GP) or other clinicians, and may also require a report from an accredited Mobility Centre. These Centres assess a person's physical and cognitive abilities thought to be necessary for safe driving and provide a recommendation. However, the large number of drivers sustaining a stroke and the associated cost of assessment mean it is impractical to refer all drivers to a Mobility Centre.

Fitness to drive is most likely to be discussed in a clinical setting as part of the rehabilitation process, and it is in this context that a simple, cost-effective, and valid assessment of fitness to drive is most needed. This information could then be used by clinicians in their report to the DVLA and help them advise the client (e.g. to cease driving whilst they await a DVLA decision).

The Stroke Drivers Screening Assessment (SDSA) was developed by researchers at Nottingham University as a tool to assess fitness to drive (Nouri et al., 1987; Nouri and Lincoln, 1992). The SDSA toolkit is sold commercially by Nottingham Rehab Ltd. It aims to help identify those post-stroke drivers with cognitive problems that may affect their ability to drive safely.

1.1 Description of the SDSA

The SDSA has four component tests:

- dot cancellation test
- square matrix direction test
- square matrix compass test
- road sign recognition test

Dot cancellation test

The dot cancellation test is designed to measure visual scanning, attention and concentration. Clients are asked to cross out all 4-dot groups on a page containing rows of 3-dot, 4-dot and 5-dot groups. They are told that they will be timed during this task, but not that the maximum time allowed is 15 minutes. Time taken to complete the test is recorded by the assessor, the number of *false positives* (incorrectly crossed out groups of 3-dot and 5-dot groups) and *errors* (4-dot groups not crossed out)[1]. The time taken to complete the task, and the number of false positives, are used in the final pass and fail equations. The number of errors is not used in these equations.

Square matrix directions test

The square matrix directions test is not included in the final pass or fail equations but is included as a 'practice test'. It is a reasoning task which has four large direction arrows and four small direction arrows. The large arrows correspond to the direction of a 'lorry' and the small arrows correspond to the direction of a 'car'. The arrow direction cards are placed along two sides of a 16 square matrix in front of the client. Cards displaying a car and a lorry facing different directions are placed in front of the client in a pile. The client is asked to position these cards onto the matrix in the correct position. One point is awarded for each correctly placed vehicle, i.e. a maximum of two per card, the maximum score being 32. If more than one card is placed in the same square of the matrix, each card is scored if correctly placed (although this is not explicitly stated in the instructions). The time limit for this test is 5 minutes. Clients are told to stop the task when the time limit is reached using a statement such as "that's fine, you can stop now".

[1] False positives are errors too, but we use the terminology from the SDSA manual.

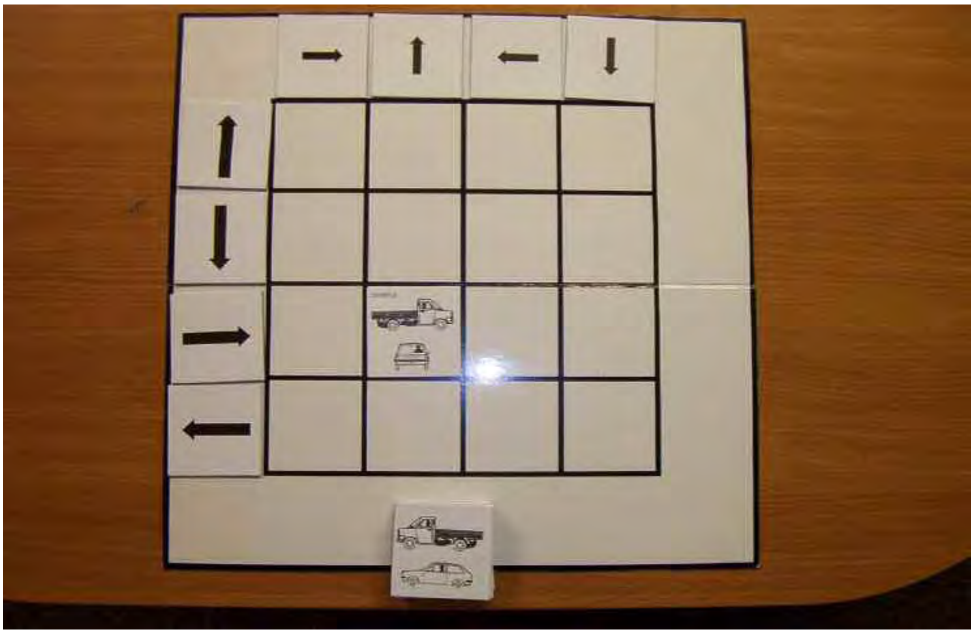


Figure 1: Square matrix directions test layout

(Photo: Lincoln et al., 2004)

Square matrix compass test

The square matrix compass test uses the 16 square matrix with ‘compass cards’ replacing the arrow cards. These cards depict arms of a compass, the black arm indicating the direction of travel. This time the cards placed in front of the client display two vehicles travelling in different directions away from the centre of a roundabout. The clients are asked to place the card in the square that matches the direction of the cars with the compass arrows. They are told that there are more cards than available spaces and that therefore some cards do not fit in the matrix. The time limit for this test is 5 minutes and it is scored in the same way as the square matrix direction test. Clients are told to stop the task when the time limit is reached using a statement such as “that’s fine, you can stop now”.

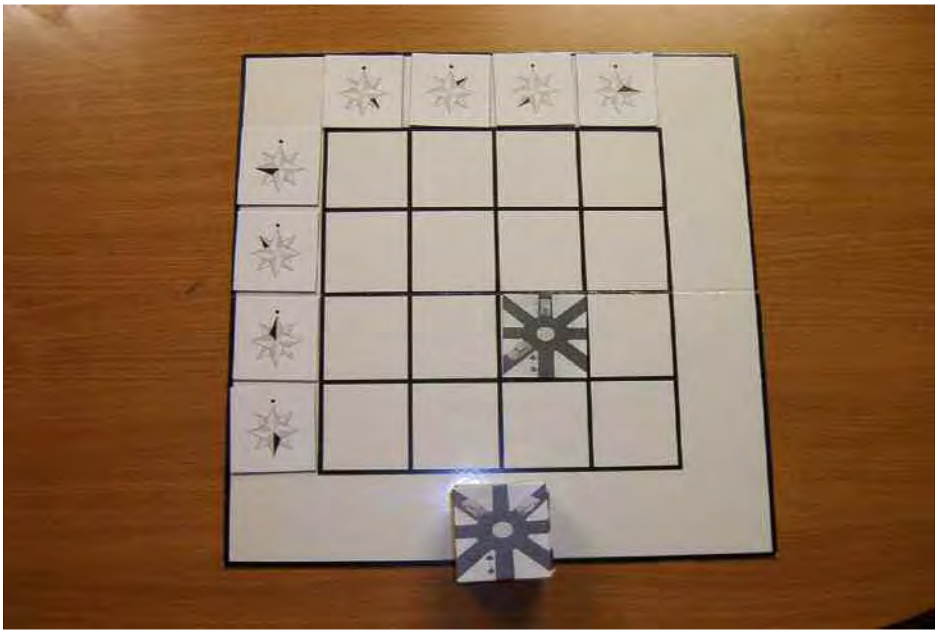


Figure 2: Square matrices compass test layout
(Photo: Lincoln et al., 2004)

Road sign recognition test

In the road sign recognition test, the client is asked to match two sets of cards for this test. One set (of 20 cards) shows various road signs, and the other set (of 12 cards) shows road traffic situations. On each road traffic situation card, the shape of the missing road sign is evident. Clients are asked to place the road sign card on top of the road situation card which they think it matches best, based on the information available in the scene. The diagram in the manual illustrates that the road situation cards should be placed in front of the client spaced out in a four by three arrangement, with example card and the road sign cards in a pile in front. Each correct match scores one point, with a maximum of 12 points. When more than one road sign card is placed on a road situation card, if one is correct it is still scored. The client is not informed that the time limit for this test is 3 minutes or that there are more road sign cards than situation cards. Clients are told to stop the task when the time limit is reached using a statement such as “that’s fine, you can stop now”.

Scoring

With the exception of the dot cancellation task errors and false positives calculations which are calculated after the SDSA test has been completed, each task of the SDSA is scored before the client commences the next task. The scores recorded include:

A: time in seconds taken to complete the dot cancellation task

B: errors in the dot cancellation task

C: false positives in the dot cancellation task

D: square matrix directions score

E: square matrix compass score

F: road sign recognition score

Scores for A, C, E and F are then put into a pass equation and a fail equation to produce two scores, the higher of which indicates whether the client has 'passed' or 'failed'. The following equations are used:

Pass score = (A x 0.012) + (C x 0.216) + (E x 0.409) + (F x 1.168) – 13.79

Fail score = (A x 0.017) + (C x 0.035) + (E x 0.185) + (F x 0.813) – 10.042

The SDSA instruction manual states that if the client has a higher value for the pass equation than for the fail equation (referred to as a 'pass' result in this report) his or her "cognitive abilities are such that driving is feasible but physical and mental aspects will need to be checked by a medical practitioner or through a Disabled Driving Centre [Mobility Centre]" (Nouri and Lincoln, undated). If the value of the fail equation is higher (a 'fail' result) the client should be advised not to drive but may be reassessed in 3-4 months, to minimise practice effects. The manual suggests that clients who are 'borderline' are referred to a Mobility Centre for further assessment.

1.2 Studies examining reliability and validity of SDSA

A screening tool needs to be both valid and reliable. The costs of making the wrong judgement are potentially high, particularly for the individual. If individuals are judged unfit to drive when they are fit, the loss of driving may result in social exclusion through loss of independence and social activities, and depression (Legh-Smith et al., 1986). If they are judged fit to drive when they are not, their safety and that of other road users may be at risk.

The validity and reliability of SDSA has been investigated in several studies by the tool's developers.

Content validity relates to whether the tool includes items that measure the cognitive/perceptual abilities impaired by stroke and relevant to driving. During the development of the SDSA, Nouri and Lincoln (1992) indicated that the Dot cancellation task was selected for the battery as a measure of attention. The task measures concentration, selective attention and unilateral inattention. A recent investigation of the content validity of the Square Matrices and Road Sign Recognition test by Radford et al. (2004) has suggested these tests both measure attention, non-verbal reasoning and spatial abilities.

The **Predictive validity** of a screening or diagnostic test is usually considered by measuring its accuracy, sensitivity, specificity and predictive values (see section 4.4.5 for definitions of these performance indices).

The ability of the SDSA to predict the result of an on-road driving assessment has been investigated by Nouri and Lincoln (1992). The road test was conducted in an automatic vehicle that was fitted with control adaptations and was assessed by one of two driving instructors. The route covered side streets, busy shopping areas, large junctions, dual carriageways and large main roads. The driver was required to complete a number of manoeuvres including parking at the side of the road, emergency stop, three-point turn, reversing around a corner and merging with moving traffic. The instructor rated the driver on 26 items, scoring either correct or fault, and graded him or her into one of three groups, pass, borderline or fail based upon their overall subjective impression of their performance on the road. Drivers the instructor considered unsafe were graded as a fail and the road test was terminated.

The study derived predictive equations from the test results of 45 subjects that correctly classified 82% of cases into the correct pass and fail road test grades. The equations were then applied to a sample of 34 subjects and found to correctly classify 79% of subjects (Nouri and Lincoln, 1992). It is not clear from the description of the analysis whether the road assessment's borderline group was combined with the pass or fail group. Sensitivity, specificity and predictive values were not reported in this study. Age and time since onset (stroke) were not found to predict road test outcome. The mean age of the sample was 61.1 years (SD 14.1) and the mean time since onset was 33.1 weeks (SD 40.7).

In a subsequent study, Nouri and Lincoln (1993) found that the SDSA correctly predicted the road test performance of 81% of patients compared to 56% correctly predicted by GPs. The SDSA results were based upon a small sample of 27 patients and a control group, assessed by GPs, of 25 patients. Sensitivity of the SDSA for detecting *passes* in the experimental group was 89% and specificity was 75%. From the published results we calculate that the Positive Predictive Value (PPV) of predicting a *fail* was 89%, meaning that an estimated 11% of those who fail the SDSA would pass the on-road assessment. The mean age of the sample was 58.8 years and the mean time since stroke was 44.4 weeks.

A small study of 21 stroke drivers by Brown (2000) found the SDSA correctly predicted on-road performance in 76% of cases and had a specificity of 73% and sensitivity of 80%. The study used a MAVIS driving assessment to assess driving performance. The assessment is the same as that described in section 3.5.6 except it included a more extensive off-road session, incorporating the manoeuvres carried out on quiet roads in the initial part of the assessment described in section 3.5.6. The mean age of the sample was 64.95 (SD 6.47) and mean time since stroke was 29.9 months (SD 35.37). Time since stroke was and age was not found to be significantly correlated with the SDSA outcome. The study suggested that square matrix compass test may be a more accurate predictor of driving performance as it alone correctly grouped 85.7% of participants.

A larger sample of 97 stroke drivers has been used to validate a Nordic version of the SDSA (Lundberg et al., 2003). This version was modified for right-hand traffic and used Swedish road signs. The road test was conducted by one of two inspectors from the Swedish National Road Administration on a set route which included a range of road types. Most of the patients drove their own vehicle. The inspectors evaluated driving performance using official criteria into pass, fail and borderline pass groups. The latter group were drivers who were deemed to be able to benefit from training. The study found that, when the borderline passes were grouped with the fail group, 67% of drivers were correctly classified using the British predictive equations, which were therefore modified. This resulted in 78% of the drivers being correctly classified but specificity was greater than sensitivity (specificity 100%, sensitivity 36%). Age was found to be significantly related to outcome of the driving test. The mean age of the sample was 63.0 (SD 12.45) and mean time since onset was 1.1 years (SD 1.45).

Test-retest reliability of the SDSA was investigated by Lincoln and Fanthome (1994). The study compared the results of the SDSA after a six week interval for a sample of 36 patients. None of the patients had been referred for assessment of their driving ability. The results showed significant improvement in the individual test components but none of the patients improved from a ‘fail’ to a ‘pass’ category as a result of practice. The scores for each component of the SDSA were highly correlated between first and second administration, with the exception of the dot cancellation time.

TRL and the University of Leeds have been commissioned by the Department for Transport to investigate further the reliability and validity of the SDSA.

2 Aims and objectives

The aims of the project were:

- To establish the extent to which the SDSA is currently being used across the United Kingdom, considering also the accuracy of use and the advice given based on the results of the test.
- To examine the validity and reliability of the SDSA on a sample of people who wish to drive following a stroke.

The specific objectives were to obtain:

- Evidence of whether the test is easy to administer and score.
- Evidence of whether, if the test is applied by more than one assessor, the same outcome is reached.
- Evidence of whether the scoring procedure and the interpretation of the results are accurate and rigorous.
- Evidence of whether, if the test is applied on two successive occasions, the same outcome (pass or fail) is obtained, and whether any learning effects arising from repeated testing with feedback result in a change in test outcome.
- Evidence of whether the test correctly classifies drivers into safe and unsafe based on their performance on a road test.

The project is described in the following four sections. First, (Section 3) there is a detailed account of the methodology. The next section (4) presents the results of the study. Finally, (Section 5) the results are discussed in relation to the aims and objectives of the project, how they relate to previous research, its limitations, and implications for further research.

3 Methodology

3.1 Overview

The project consisted of four studies:

- A **survey of SDSA users** to establish the extent to which the SDSA is currently being used across the United Kingdom, consider the ease of its use, and investigate the advice given based on the results of the test.
- An **inter-rater reliability study** to investigate whether different assessors reach the same outcome and whether the scoring

procedure and the interpretation of the results are accurate and rigorous.

- A **test-retest reliability study** to investigate whether a person tested on two separate occasions will obtain the same result and whether any possible learning effects resulting from repeated administration of the test are likely to lead to a change in the outcome (from fail to pass).
- A **validity study** to investigate the criterion-related validity (predictive validity and concurrent validity) and face validity of the SDSA.

  Criterion-related validity compares test results with a criterion/standard where the criterion is a measure of what the test is supposed to predict (i.e. 'unsafe' driving in the case of the SDSA). Accident liability is rarely used in studies of test validity as a criterion for safe driving because it would not be ethical to allow drivers who are believed to be unsafe to drive in order to continue to assess whether they are more likely to have an accident. It is also impractical as accidents are rare events and sample sizes tend to be small when sampling clinical populations. If accident liability cannot be used, the next best criterion is actual driving performance in real traffic conditions.

  There are two types of criterion-related validity: predictive validity and concurrent validity.

  Predictive validity refers to the ability of the test to predict future performance, such as performance in the driving assessment.

  Concurrent validity relates to the extent to which the results of the test under investigation correlate with the current results of other tests known to tap the variable of interest (safe or unsafe driving).

  Face validity refers to the extent to which the test seems, on the face of it, to be a valid measuring tool. Face validity is an important factor for clients whose driving privileges are at stake. An on-road driving assessment has high face validity whereas the SDSA is likely to have less face validity.

3.2 Survey of SDSA users

This study consisted of a postal survey of healthcare professionals and researchers and telephone interviews with a sub-sample of survey respondents to gain more detailed information about issues raised in the survey.

3.2.1 Survey sample

Resumption of driving is seen as an important milestone in recovery by people who have sustained a stroke. Occupational Therapists (OTs) and others working in rehabilitation services are therefore often asked by clients for advice about returning to drive.

To determine the extent to which the SDSA is currently being used in the United Kingdom it was necessary to approach a large number of healthcare professionals who were potential users of the SDSA.

A self-completion questionnaire was distributed by two professional organisations to their members. 750 questionnaires were issued by the National Association of Neurological Occupational Therapists (NANOT) and 325 by the British Society of Rehabilitation Medicine (BSRM). Twenty questionnaires were also sent to people known to the researchers to be using the SDSA. These people were recruited from a variety of sources including the British Psychological Society (BPS) website, SAGE ("Safer driving with Age" mobility agency) and the Forum of Mobility Centres.

3.2.2 Questionnaire

The postal questionnaire was designed to elicit information about the methods used to assess fitness to drive of people following a stroke and more specifically about the use of the SDSA. The questionnaire covered the following issues:

- identification of the range of tests used to assess the fitness to drive of people who have sustained a stroke;
- who administers the SDSA and their level of training and experience;
- ease of administering, scoring and interpretation of the SDSA tests;
- users' views on its validity and reliability;
- comments received from clients about the SDSA (e.g. face validity issues);
- the extent to which the SDSA is used and relied upon in advising clients about driving;
- the advice given to clients regarding fitness to drive based on the SDSA score;
- the assessment process (how stroke clients get referred for an assessment, who receives the advice and how it is acted upon).

3.2.3 Sample of interviewees

Twenty survey respondents were interviewed by telephone to gain more detail about key issues. The sample included some of the postal survey respondents in each of the following categories:

- Respondents who used the SDSA as part of a test battery.
- Respondents who reported experiencing difficulties in administering the SDSA.
- Respondents who reported clients had difficulty completing and/or understanding the tests.
- Respondents who reported clients had difficulty accepting the result of the SDSA.

3.3 Inter-rater reliability and interpretation issues

To investigate inter-rater reliability and interpretation issues a video showing three SDSA assessments carried out by a trained user was sent to current SDSA users. Users were asked to point out any difference between the administration procedure shown on the tape and their own practice. They were also asked to score and interpret the results of the tests shown on the video and say what their final advice to each client in the video would have been.

3.3.1 Recruitment of raters for the inter-rater reliability study

A number of current SDSA users were identified from the survey of users described above. These users had indicated that they were willing to take part in further research and supplied contact details. Each SDSA user was contacted by telephone and invited to take part in the study. The users were also invited to distribute additional copies of the study materials to colleagues who currently use the SDSA. All users participating in the study were informed that the data and information supplied by them would be confidential and anonymous.

3.3.2 Study procedure

Raters were shown a video of three stroke clients using the SDSA. The video was specially produced for this study using a professional studio at the television department of the University of Leeds. It depicted the SDSA being administered to three stroke patients by an expert user, Professor Nadina Lincoln (a developer of the SDSA). The stroke patients were recruited from in-patients at the Chapel Allerton Hospital in Leeds and had given informed consent to taking part. A standardised script was developed to ensure that the filming of the administration of the SDSA to patients was consistent and that the cameraman knew which aspects to film at different points.

Raters were asked to score each client as they normally would and provide a result and their recommendation for the client. An instruction sheet was provided with information on how to complete the task, such as pausing the tape to score the square matrices and road sign recognition task.

The raters also completed a questionnaire, which sought information on the problems encountered with the scoring or interpretation of each test. They also observed differences between their own administration of the SDSA and that shown on the video.

The process was expected to take approximately one hour to complete. A £10 gift voucher was given to those who returned their materials.

Reminder letters were sent to those who had not returned their study materials after 8 weeks. A total of 65 sets of study materials were distributed to SDSA users; 28 were returned.

3.4 Test-retest reliability and learning effects

To investigate test-retest reliability and learning effects, experimental participants (i.e. people who were thinking of returning to driving following a stroke) taking part in the validity study were asked to participate in the test-retest and training effects study. A total of six people volunteered and completed the study. To minimise the possibility that any observed effect was due to the natural improvement in the stroke symptoms, the second administration of the SDSA was within 6 weeks of the first administration.

After this, the participants completed the SDSA three further times within a week and received feedback on their performance during the first two of these sessions: a trained assessor pointed out the errors made at the end of each test. No additional information about each test (such as that it was timed or there were more cards than situations in the road sign recognition test) was given, and strategies on how to complete the test were not discussed. This procedure enabled us to examine the extent to which a person who wishes to pass the SDSA might be able to improve his or her performance by practicing the SDSA with someone who is able to give feedback on errors.

This study was completed at the home of the participant, so due to practical reasons, the sample was limited to people who lived within 30 miles of TRL.

3.5 Validity study

To investigate whether the SDSA can distinguish between drivers who are safe and unsafe an experimental validity study was undertaken. The criterion used was actual driving performance in real traffic as measured by whether the participant passed or failed a driving assessment. The study used a ‘double blind’ methodology, where the assessors of the SDSA and driving assessment were not told the result of the other assessment or the condition of the participant (experimental or control).

3.5.1 Procedure

The project involved collaboration between TRL, the University of Leeds and DfT’s Mobility Advice and Vehicle Information Service (MAVIS). The study was based at TRL in Crowthorne, Berkshire.

Experimental participants were recruited through the clinics of four hospitals, advertisements placed in the Stroke Association’s and Different Stroke’s newsletters, local stroke clubs, local press coverage, and personal contacts of the research team. Participants were selected according to specified inclusion criteria (see section 3.5.2). If they met the inclusion criteria, participants visited TRL and undertook the SDSA and on-road driving assessment. Some participants also completed a Useful Field of View (UFOV) test. Further information on these tasks is given in sections 3.5.4 to 3.5.6.

Control participants were recruited from the TRL database of volunteers. This database holds details of drivers who are willing to take part in research trials. Volunteers were not recruited as controls if they had had a stroke. The control participants also visited TRL and undertook the SDSA. Some participants also completed the UFOV and on-road driving assessment.

The progress of a participant through the project is shown in Figure 3.

3.5.2 Participants

Experimental participants were selected according to the following criteria:

- were at least one month post-stroke
- held a full driving licence
- had driven for at least two years prior to their stroke
- drove at least 30 miles per week prior to their stroke

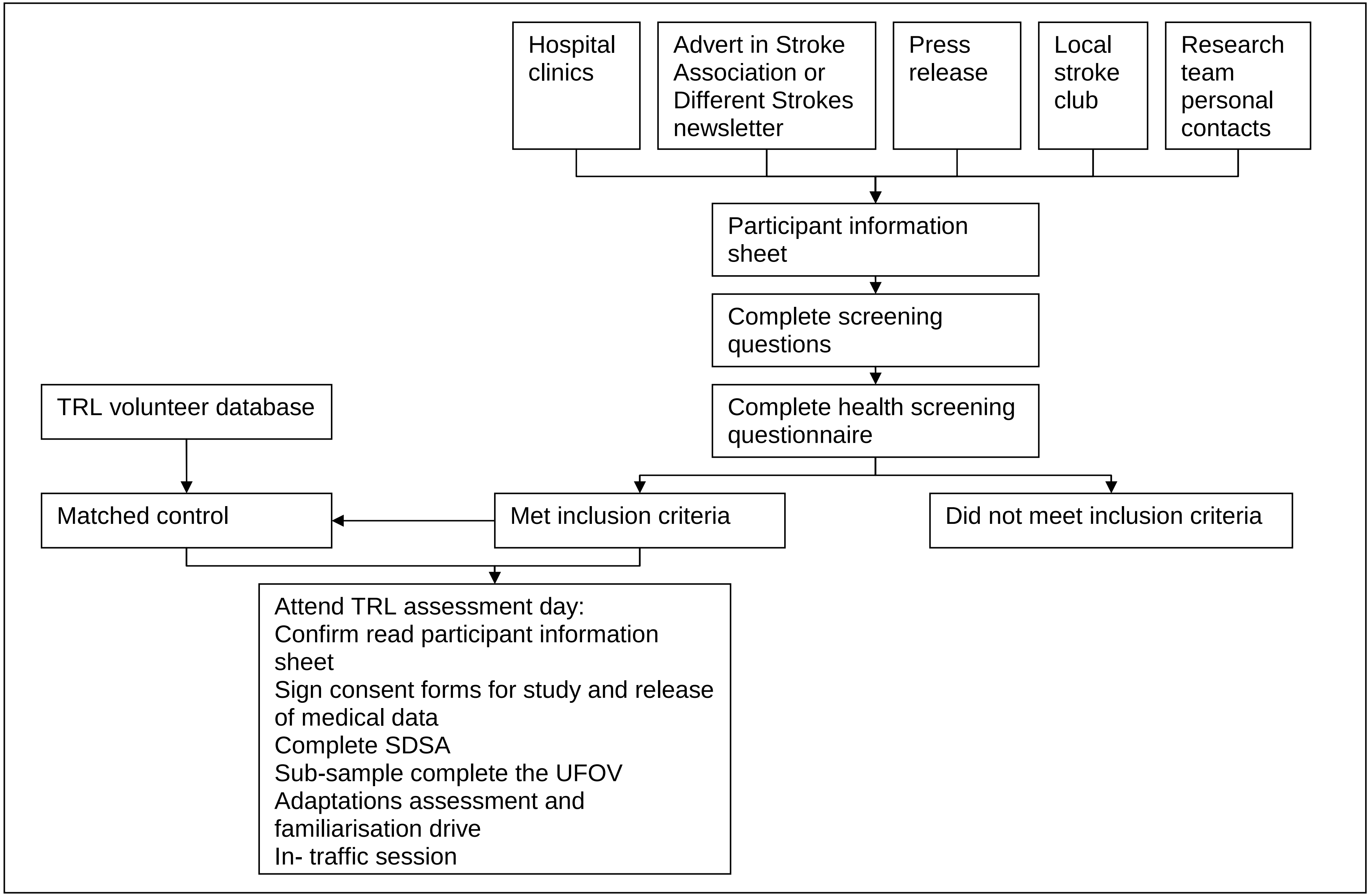
Prospective participants were excluded from the study if there were other prescribed disabilities (i.e. a disability or health condition reportable to DVLA Medical Branch). This was established using a health screening questionnaire that was completed by the prospective participant and returned to TRL by post in a pre-paid envelope. Prospective participants were excluded if they had already received a decision from DVLA on their fitness to drive. Prospective participants who lived more than 100 miles from TRL were also excluded.

Participants needed to be in possession of a valid driving licence and produce it on the day of the assessment in order to be taken in-traffic. This was a requirement of the insurance, to ensure participants' licences were valid and had not been revoked. Experimental participants who had not retained their licence whilst they awaited a decision from DVLA were, therefore, excluded from the study.

The control sample were recruited from the TRL volunteer database and matched experimental participants, where possible, for age, gender and driving experience (years driving).

Those participants who undertook the neuro-psychological tests (SDSA and UFOV) and driving assessment received £25 plus travel expenses. Participants who only undertook the neuro-psychological tests received £15.

Figure 3. Progress of participant through research study



3.5.3 Medical information

Experimental participants were asked to give consent for details of their stroke relevant to the project to be made available to the research team. Their doctor was then contacted and asked to complete a questionnaire about their last stroke before the assessment. The questionnaire sought information on:

- Side of lesion
- Type of stroke
- Visual disturbances
- Motor disturbances
- Sensory disturbances
- Higher Cognitive dysfunction
- Incoordination (Ataxia)

3.5.4 SDSA

The SDSA was administered by a trained assessor on the same day as the other assessments. The assessor was 'blind' to the results of the driving assessment and participant's condition (experimental or control).

3.5.5 Useful field of view

The Visual Attention Analyser (Useful Field of View test, UFOV) was also administered to a sub-sample of participants. The UFOV measures the total visual area in which information can be acquired without eye and head movements (Ball et al., 1988) and has been shown to be predictive of accident involvement in older drivers (Ball et al., 1993). However, it has not been used to predict the accident involvement or actual in-traffic driving performance of drivers following a stroke, and so its validity for this purpose is not known. It was included in this study partly because in the absence of any other test known to be predictive of safe driving for drivers following a stroke it could provide at least some indication of concurrent validity of the SDSA. In addition, the data collected by the study would provide an indication of whether the UFOV itself has potential to identify those stroke drivers who would fail the MAVIS on-road assessment.

The UFOV is a computer based test that measures speed of processing for increasingly complex tasks. The first sub-test presents a target (car or truck) in a central box, which the participant must identify. In the second sub-test, which measures divided attention, the participant must identify the central target (car or truck) and a peripheral target (a car). In the third sub-test, the participant must identify the central target and a peripheral target embedded amongst 47 triangles. The last sub-test measures selective attention. Each sub-test presents a score in milliseconds (msecs). If the score is greater than 500 for first or second sub-test, subsequent sub-test(s) are not administered and a default score of >500 is given. If the test is aborted a value of 0 is assigned. Based upon various combinations of the sub-test scores, the UFOV categorises participants into one of five different risk categories:

1. Very low risk
2. Low risk
3. Low to moderate risk
4. Moderate to high risk
5. High risk and very high risk

The test was administered by the same assessor as the SDSA, who was ‘blind’ to the results of the driving assessment and participant’s condition (experimental or control).

3.5.6 On-road driving assessment

The on-road driving assessment was conducted by either a MAVIS driving advisor or one of five Approved Driving Instructors trained by the MAVIS driving advisor to give adaptation advice and carry out an on-road driving assessment for drivers who may be impaired as a result of a stroke. As part of the training process the driving assessors all rode in the same vehicle to observe the same 3 drivers so as to promote consistency between assessors. Refresher training was also provided to ensure that scoring remained consistent.

The driving assessment consisted of two stages:

- an adaptations assessment and off-road familiarisation drive; and
- an in-traffic session.

3.5.6.1 Adaptations assessment and off-road familiarisation drive

Prior to the in-traffic session, the drivers’ need for adaptations, such as infra-red controls or left-foot accelerator was assessed by a driving instructor. An automatic transmission vehicle was used for all the sessions. Participants were given an hour long session off-road, around TRL’s small roads system, to become familiar with driving an automatic vehicle with or without adaptations. If the driving instructor felt the driver was unsafe to continue to the in-traffic session, the trial was terminated following the familiarisation session and the participant was assigned a ‘fail’ result.

3.5.6.2 In-traffic session

To prevent participants becoming too fatigued they had a short break between the familiarisation session and the in-traffic assessment.

A standard route was used, which was based on that used as part of a MAVIS driving assessment and also used the same scoring procedure. Although the MAVIS on-road driving assessment has not been validated, it has been used in previous fitness to drive studies (Christie et al., 2001; Grayson et al., unpublished; Brown, 2000).

The driving assessor was ‘blind’ to the results of the SDSA and UFOV and had not carried out the adaptations assessment or familiarisation drive so was not familiar with a participant's condition or performance. Few of the drivers required adaptations so it was not obvious whether the driver was in the control or experimental group. The driving assessor was ‘blind’ to their condition.

All the participants drove the same automatic vehicle.

Initially, the driver was asked to undertake a series of exercises on quiet roads within the TRL site, as described in Table 1. These manoeuvres were believed by MAVIS to assess the driver’s motor co-ordination, spatial ability, orientation and information processing.

Table 1. Quiet road manoeuvres

Short description of exercise	*Elements thought by MAVIS to be tested in individual exercises*
Ability to drive through a coned circuit	Motor co-ordination Spatial ability Information processing
Actions, observations and control of the car when reversing around a corner to the left	Spatial ability Orientation
Ability to understand instructions, actions, observations and control of the car when reversing into a parking space on the right	Spatial ability Orientation

If the assessor felt that the participant had performed these exercises safely, the assessment moved out into the traffic conditions of the Sandhurst, Crowthorne and Bracknell areas. The in-traffic session could be terminated early if the assessor felt it was unsafe or the participant did not want to continue. In these cases the participant failed the driving assessment.

The in-traffic tasks were thought by MAVIS to assess skills such as decision making, information processing in high workloads, selective and divided attention, positioning at 60 mph, and interaction with other road users. Table 2 lists the scored exercises and elements thought to be tested. A total of 19 exercises were scored in total.

Table 2. In-traffic exercises

Short description of exercise	*Elements considered for score*	*Elements thought by MAVIS to be tested in individual exercises*
Behaviour at five roundabouts	Regulation of speed on approach to roundabout Position on approach to, while on and when leaving roundabout Observations and actions at roundabout	Decision making
Count out loud slowly to 30 at roundabout	Effects on counting on driving Effects of driving on counting	Divided attention
Behaviour at four mini roundabouts	Speed of approach, position and observations and actions	Decision making
Detect three given signs	Observations and indication of three ‘elderly people’ signs	Selective attention
High Street	Observations at pedestrian crossings, awareness of traffic-calming system and actions in relation to other road users	Interaction with other road users and road features
Drive at 60 mph	Positioning while driving at 60 mph	Spatial ability, ability to set and maintain speed
Behaviour when making two specific lane changes	Observations of other traffic, indication and timing of lane change	Interaction with other road users

Following signs to M4 over a series of 4 roundabouts	Ability to read and follow signs Approach and observation at roundabouts Interaction and awareness of other road users	Information processing under high workload conditions

Each exercise was given a score on a scale 1 to 4. A score of 4 indicates definitely safe, a score of 3 indicates probably safe, a score of 2 indicates probably unsafe and a score of 1 indicates definitely unsafe. Using these scores a mean score was calculated for the main skills thought to be assessed:

- decision making (9 scores);
- divided attention (1 score);
- interaction with other road users (3 scores);
- selective attention (1 score);
- spatial ability (1 score); and
- information processing under high workload conditions (4 scores).

A total driving score was also calculated based upon the mean of the 19 in-traffic scored exercises. This variable was used when investigating the correlations between variables and differences between groups.

The assessor also gave an overall judgement on whether the driver was at a greater risk than normal of an accident (yes or no). As is the case with the standard MAVIS driving assessment, this judgement was based upon patterns in the scores and comments rather than the total score. The patterns are based upon identifying the lowest score in each skill group and considering whether the pattern of scores highlights a cognitive deficit. The scores for the manoeuvres carried out in the initial part of the assessment are also considered. This risk variable was used to group drivers into pass and fail groups. Those considered at greater risk than normal of an accident were classified as fails and those not a greater risk as passes. This variable was used in all analyses which examined differences between driving assessment pass and fail groups.

The in-traffic assessment lasted approximately 45 minutes.

The majority of on-road driving assessments were conducted by one driving assessor (44), with the other assessors carrying out between one and 13 on-road driving assessments.

3.5.7 Ethical considerations

The DVLA Medical Branch was consulted in the development of an ethics protocol for this study. In the UK, it is the duty of the driving licence holder to notify DVLA of a relevant or prospective disability such as a stroke. The experimental participants should already have informed DVLA of their condition but not all will have done so for various reasons. The participant information sheet advised of their need to inform DVLA.

The research study was conducted independently of the DVLA process determining fitness to drive and the results were not disclosed to anyone, including the participant, their doctor or the DVLA.

The project methodology was approved by the South West Multi-Centre Research Ethics Committee and received locality approval from the Local Research Ethics Committee of referring clinics.

Prospective participants were informed that:

- Participation was voluntary.
- Information from the assessments would benefit the rehabilitation process for drivers following a stroke in the future.
- Whether or not they agreed to participate would not affect their treatment.
- The outcome of all the assessments would be anonymous and confidential.
- No-one, including them, their doctor or the DVLA, would be told outcome of the assessment.

Participants were given a copy of the Stroke Association's leaflet on returning to drive after a stroke at the end of their assessment day.

3.5.8 Face Validity

The extent to which clients questioned the face validity of the SDSA was evaluated as part of the survey of SDSA users. Anecdotal comments made by participants were also collected during the validity study.

4 Results

4.1 Survey of SDSA users

The main findings from the survey are reported in this section. More detailed information can be found in Sentinella et al., unpublished a.

4.1.1 Response rate

Identifying users of the SDSA proved problematic and a low response rate of 19.5% (213 respondents) was achieved from distributing questionnaires to the two largest professional organisations most likely to contain potential users of the SDSA and placing an advert on the British Psychological Society website. However, it is difficult to estimate an accurate response rate because respondents were invited to copy and distribute the questionnaire to others, potentially increasing the number of questionnaires issued. It is also possible that some people were members of more than one of the professional organisations sampled and received multiple copies of the questionnaire. The low response rate was not unexpected, as the questionnaire was unlikely to be completed by people who had not heard of the SDSA or did not use it.

4.1.2 SDSA users

Sixty-three of the 213 respondents were currently using the SDSA and a further 24 had used it in the past. If it is assumed there is no duplication of professional membership, the results suggest that at least 6% of the members of NANOT and BSRM are currently using the tool in the UK.

Ninety-four per cent of current SDSA users were occupational therapists (59 respondents). Other users included four Consultants in rehabilitative medicine.

4.1.3 Use of the SDSA in the UK

Respondents were asked for which type of clients they used the SDSA (n=63):

- 62 (98%) reported they used the SDSA for people who have sustained a stroke;
- 22 (35%) reported also using the SDSA for people with a head injury;
- 7 (11%) reported using the SDSA for other clients including those with Multiple Sclerosis, Parkinson’s disease or dementia. One respondent

also reported using the SDSA with clients who had had viral meningitis.

Thirty respondents reported that they used other tests as well as the SDSA as part of their assessment process for clients wanting to return to driving. The tests used are detailed in Table 3 below.

Table 3: Tests used in conjunction with the SDSA (multiple responses possible)

Test used with SDSA	*Number of respondents citing each test*
Chessington Occupational Therapy Neurological Assessment Battery	10
Rivermead Perceptual Assessment Battery	9
Behavioural Assessment of the Dysexecutive Syndrome	6
Rivermead Behavioural Memory Test	5
Middlesex Elderly Assessment of Mental State	5
Behavioural Inattention Test	3
Loewenstein Occupational Therapy Cognitive Assessment	2
Mini Mental State Examination	2
Cognitive Assessment of Minnesota	1
Total number of respondents using one or more tests in addition to the SDSA	30

When a sub-sample of survey respondents were probed about this issue during the telephone interviews it was found that the SDSA was not being used as part of a strict assessment battery. The combination of tests used, either as screening tests before the SDSA or as additional tests alongside the SDSA, was entirely dependent on the individual client and down to the OT’s judgement.

4.1.4 Opinions on administering, interpreting and scoring the SDSA

In terms of administering, interpreting and scoring of the component tests, the ‘square matrix compass’ test was considered the most difficult to use and the 'dot cancellation' test was seen as the easiest to administer and interpret. Scoring this latter test was difficult but the problem could be easily solved by producing a scoring overlay. Many of the SDSA users have done this themselves.

A difficulty was reported with the compass cards on the 'square matrix compass' test, in that only half of the compass point is shaded, which is inconsistent with the pictures in the instruction manual. This shading confuses clients and users about the compass heading it is supposed to depict.

It was reported by SDSA users that with the 'square matrices' tests there were so many different cards on the board they were easily knocked out of position by mistake and this could alter the direction of the arrow or compass. It was suggested that the cards could be fixed to the square matrix board with 'Velcro' to prevent this happening.

It was also reported that there were no scoring criteria for the 'road sign recognition' test and the SDSA user had to use their own knowledge to score the clients' responses. This implies that all SDSA users are driving licence holders and/or have a good knowledge of the Highway Code, which is apparently not always the case. Providing a scoring guide in the instruction manual could easily rectify this.

A number of comments were made about the types of difficulty clients encountered using the test such as unfamiliarity with the road signs in the 'road sign recognition' test or differentiating between the depiction of the front and rear view of a car in 'square matrix direction' test.

Some respondents felt there was a need for more information to be included in the manual such as an explanation of the formulae used, the theory and research behind the tests and more detailed instructions.

The nature of standardised testing demands that the instructions must be given as they are stated in the manual, are only repeated once and cannot be elaborated upon. But many SDSA users reported that the instructions were not sufficient to explain the tasks to the client and it was believed that many were failing tasks because they simply did not understand what it was they were supposed to do. This actually caused some SDSA users to further explain the tasks involved, thus deviating from the written instructions given in the manual and undermining the SDSA's status as a standardised test.

Users expressed concern that some of the tests have to be marked in front of the client. This requires the client to wait whilst the marking takes place, and may cause them some anxiety.

In terms of interpretation of the equation scores, some SDSA users reported uneasiness about classifying scores as a ‘pass’ or ‘fail’, especially if the score was ‘borderline’. This uneasiness was often the case when they felt that the score did not reflect the clients’ true cognitive ability. It was suggested that advice be provided on a ‘borderline’ category. This problem was confounded by the fact that the SDSA users were normally referring clients back to their GPs who have little knowledge of the SDSA.

Some of these problems associated with the administering, interpreting and scoring the SDSA sub-tasks could be due to the very low number of SDSA users having actually received training in its use or their lack of experience in administering the tool.

4.2 Inter-rater reliability and interpretation of SDSA scores

Inter-rater reliability of the SDSA was assessed using the intra-class correlation coefficient (ICC). The analyses were based upon the SDSA scores for 3 clients, each scored by 28 different raters: a total of 84 SDSA scores.

Table 4 illustrates that the ICCs are all high (ranging from 0.82 to 1.0), suggesting a very high level of agreement between raters for each of the components of the SDSA.

Table 4: Inter-rater reliability for SDSA components (n = 84)

SDSA Component	ICC
A: Time for dot cancellation (secs)	1.00
B: Errors in dot cancellation	0.98
C: False positives in dot cancellation	0.86
D: Square Matrix Directions score	0.98
E: Square Matrix Compass score	0.95
F: Road sign recognition score	0.99
Pass equation score	0.82
Fail equation score	0.85

Table 5 shows the scores for each client assigned by the expert user shown in the video.

Table 5. SDSA scores assigned by the expert user for each client shown in the video.

SDSA Component	Client 1	Client 2	Client 3
A: Time for dot cancellation (secs)	204	900	467
B: Errors in dot cancellation	113	70	43
C: False positives in dot cancellation	28	26	0
D: Square Matrix Directions score	21	6	32
E: Square Matrix Compass score	27	12	16
F: Road sign recognition score	9	1	9
Pass equation score	16.26	8.70	8.87
Fail equation score	6.72	9.21	8.17
Outcome score (pass – fail score)	9.54	-0.51	0.07
Outcome (based on whether the pass or fail equation score is greater)	Pass	Fail	Pass

The SDSA outcome should be based upon whether the ‘pass equation’ score or ‘fail equation’ score is greater. Therefore, based upon the scores assigned by the expert user shown in the video, client 1 and client 3 ‘passed’ the SDSA and client 2 ‘failed’. Table 6 shows that the outcome assigned by raters was generally consistent for client 1, most raters gave the client a ‘pass’ outcome, whilst the level of agreement between raters for clients 2 and 3 was mixed. Table 6 shows the outcome the raters would assign each client based upon their own calculations and judgements.

Table 6. Outcome assigned by raters

	Pass	Fail	Borderline	Other	Total
Client 1	25	1	0	2	28
Client 2	1	18	8	1	28
Client 3	16	2	10	0	28
Total	42	21	18	3	84

In three cases raters gave an ‘other’ outcome because they did not feel the client should not have ‘passed’ (the pass equation score was higher than the fail equation in each case). For client 2 and client 3 the difference between the pass and fail scores was small, consequently, the result given by raters was mixed. Table 6 shows that a ‘borderline’ outcome was frequently used in these cases, although the questionnaire responses highlighted that there was confusion and difference of opinion over what constituted a ‘borderline’ result or even if it existed. The outcome score (the difference between pass and fail equations) for the two clients assigned ‘borderline’ results in this study ranged from -1.11 to +1.03. Client 2 was given a borderline outcome by eight raters. In seven of the cases, based upon the raters’ own scores, client 2 had a positive outcome score and in one case the outcome score was negative. Most of the raters (18) gave client 2 a ‘fail’ outcome, which is in line with the outcome the expert users’ scores indicated. Similarly, in eight cases client 3 had a positive outcome score but the rater gave the client a borderline result, and in two cases client 3 had a negative outcome score but was assigned a borderline result. The majority of raters gave client 3 a ‘pass’ outcome, which is the outcome the expert user’s score indicated.

Concern was expressed by some raters that the outcome of the SDSA did not reflect the clients’ performance in the tasks as shown in the video, perhaps suggesting a lack of confidence in the SDSA’s ability (at least, using the current scoring system) to predict driving performance.

In the study only one rater recommended that a client in the video be told at that time he or she was unfit to drive based solely upon the SDSA results. The majority of raters recommended that all three clients shown in the video should all undergo further cognitive, physical or driving assessment, despite all the raters assigning a ‘fail’ outcome to at least one client. Two of the raters would have referred the clients with a ‘fail’ outcome to a mobility centre for driving assessment; three raters would have referred the clients with a ‘borderline’ outcome to a mobility centre and five of the raters would have referred clients with a ‘pass’ outcome to a mobility centre. On this evidence, raters do not appear to use the tool as a screening test to identify clients requiring further assessment.

4.2.1 Scoring the SDSA for clients shown in the video

Eleven of the 84 SDSA score sheets contained calculation errors made by the raters. The majority of errors were made when multiplying or summing the SDSA components in the equations. Other errors included recording the time for the dot cancellation test in minutes rather than seconds and transcribing an incorrect figure into an equation from the individual scores. In the majority of cases the errors did not affect the overall outcome (i.e. whether the pass or fail equation score was greater).

Comments made by respondents in the questionnaires revealed that it is not clear to users why the scores for dot cancellation test errors (B) and the square matrix directions test (D) are recorded on the score sheet when they are not used in the pass or fail equations.

There is no indication from the study that raters' used the scores on the individual SDSA subtasks in their interpretation of the result, although the questionnaire did not seek this information directly.

4.2.2 Observed differences in the administration of the SDSA

All the 28 raters who took part in the study were current users of the SDSA. Eleven raters reported that their own administration of the SDSA differed from that shown in the video. Some raters said they gave additional instructions or information to the client, especially if the client did not appear to understand the instructions given in the SDSA instruction manual.

Another common difference in administration was in the presentation of the cards in the road sign recognition test. Some raters commented that they placed the road sign cards in a pile in front of the client, rather than spread across the table in front of the client as shown in the video. If cards are spread across the table, clients may find it easier to select the matching sign for the road situation cards by looking at the situation card first and then the signs. If the cards are presented in a pile, this may lead the client to look at the road sign card and try to place it in the correct situation, potentially 'wasting time' as clients are not warned that there are more sign cards than situations and the test is time limited. Some raters warn clients that there are more sign cards than situations and the time allocated for each test.

When scoring the dot cancellation test one rater reported that she previously included the practice line in the timing of the dot cancellation task. However, the manual states that this should not be included.

For more detailed information on the inter-rater reliability study see Sentinella et al. (unpublished b).

4.3 Test-retest reliability and learning effects

4.3.1 Overall test-retest reliability

Only six participants volunteered for and completed this part of the project. Of these, only two failed the SDSA at first administration. All those who passed on the first occasion, and one of those who failed on the first occasion passed the SDSA on the second occasion. After two training sessions, the same overall outcome as that obtained at the second administration was maintained. The participant whose outcome changed between first and second administration had outcome scores of -0.69 and +4.16 on first and second administrations respectively. The first score was closer to a pass than that of the participant who failed on both occasions. The participant who changed SDSA outcomes was 8 months post-stroke at the time of first administration, whilst the other five participants were between 11 months and 9 years. All the participants who passed on both occasions were a year or more post-stroke.

Test-retest reliability of the individual components of the SDSA was assessed using the ICC statistic. It examines the rank order of the participants on retest. A high correlation means the participants kept a similar rank on retest and so there is a high level of agreement between the scores.

The ICCs indicate that there were varying levels of agreement in the scores between first and second administration for each SDSA components except the dot cancellation test false positives score, for which there was no agreement (ICC = 0.00) because five of the six participants scored zero on both occasions. The ICCs show reasonable levels of agreement between scores on the dot cancellation time, dot cancellation test errors and square matrix test, the ICCs ranging from 0.70 to 0.98. The ICC for square matrix compass test showed less agreement between scores (0.58), whilst the negative retest correlation for the road sign recognition test (-0.10) indicated the participants' rank order in the first test had been reversed on retest. The varying levels of agreement for the individual components of the SDSA are reflected in the equation scores.

Table 7. Test-retest reliability for SDSA components (n = 6)

SDSA Component	ICC
A: Time for dot cancellation (secs)	0.98
B: Errors in dot cancellation	0.70
C: False positives in dot cancellation	0.00
D: Square Matrix Directions score	0.75
E: Square Matrix Compass score	0.58
F: Road sign recognition score	-0.10
Pass equation score	0.26
Fail equation score	0.09
Outcome score (pass – fail score)	0.55

4.3.2 Changes in mean scores

On average the second administration of the SDSA was 22.5 days after the first administration (SD 10.5).

The paired-samples t-test was used to compare performance on mean scores for each SDSA sub-test at time 1 (first administration) and at time 2 (second administration). There was a statistically significant difference in the time to complete the Dot cancellation test (p=0.02), with the mean time taken at second administration reduced (415 versus 450 seconds). This relationship holds true both for those who passed and for those who failed SDSA at first administration and were statistically significant (p=0.008 and p=0.03 respectively (see Table 8)).

Table 8. Mean dot cancellation time at first and second administration for those who pass and who fail SDSA on first administration

	Outcome of SDSA at first administration	
	Pass (n=4)	Fail (n=2)
	Mean dot cancellation time (in seconds)	
First administration	421	508
Second administration	380	486

Although not reaching statistical significance in our small sample, the scores for the square matrix direction test and square matrix compass test showed indications of improvement, with the mean scores increasing on second administration both overall and separately for those who passed and those who failed SDSA at first administration. The p-values for these differences were p=0.11 and p=0.10 respectively.

The outcome score (i.e. the difference between the pass equation and fail equation scores) was explored in more detail. A positive outcome score indicates a ‘pass’ outcome and a negative score indicates a ‘fail’. When the mean outcome scores at first administration (time 1) and second administration (time 2) were compared using the paired-samples t-test, an improvement in mean score was found that approached statistical significance. The mean outcome score increased from 0.25 at time 1 to 2.31 at time 2, a difference that and approached statistical significance (p=0.09).

There was no statistically significant difference in outcome scores between second administration and final (fifth) administration, after two intermediate sessions in which feedback was given to the candidate. However, the difference between first administration and final administration was statistically significantly different (p=0.01). The mean outcome score increased from 0.25 at time 1 to 3.16 at time 5. The improvement in scores was largest in the pass group; the mean outcome score increasing from 1.29 at time 1 to 4.36 at time 5 as shown in Table 9. The difference between the same scores for the fail group was not statistically significant.

Table 9. Mean outcome score[1] at first, second and fifth administration for those who pass and who fail SDSA on first administration

	Outcome of SDSA at first administration	
	Pass (n=4)	Fail (n=2)
	Mean outcome score[1]	
First administration	1.29	-1.81
Second administration	3.20	0.52
Fifth administration (after training)	4.36	0.78

[1] Outcome score = difference between pass equation and fail equation scores

4.3.3 Learning effects

To examine the effects of training on SDSA sub-test scores, paired-samples t-tests were used to compare performance on each scored variable recorded during the test at time 1 and at time 5 (after two training sessions). There was a significant difference between the mean square matrix compass scores (p=0.002), with a higher score after training (26 versus 17). This relationship holds true both for those who passed and for those who failed SDSA at first administration, although the differences were not statistically significant (see Table 10).

Table 10. Mean compass score before and after training for those who pass and those who fail SDSA at first administration

	Outcome of SDSA at first administration	
	Pass (n=4)	Fail (n=2)
	Mean compass score	
First administration	22	7
After training	30	19

The dot cancellation errors score and the road sign recognition score both showed near significant effects (p=0.086 and p=0.093 respectively) in our small sample, with the mean scores showing improvement after training on both measures. This applied when all volunteers were considered together and when those who passed and those who failed SDSA at first administration were examined separately.

4.4 Validity study

4.4.1 Participants

The sample comprised 88 people, including 42 stroke participants and 46 control participants.

4.4.1.1 Experimental participants

Ninety-two people who had had a stroke were referred to the study. Of these, 24 people were ineligible for the study either because they had a prescribed disability (3 referrals), had already received a decision on their fitness to drive from the DVLA (9 referrals), or had not retained their driving licence whilst awaiting a decision from DVLA (10 referrals). Two people lived outside the study area. In addition to the 24 ineligibles, three people decided not to take part in the study and eight did not return their health screening questionnaire, despite being sent reminders.

Of the 57 who remained, four people subsequently withdrew from the study due to ill-health, one person was too anxious to take part, another person had no-one to accompany her and did not want to come alone, one person needed to care for his sick wife, two people had availability problems and three felt it was too far to travel. A further three people postponed their assessment date and it could not be rescheduled before the end of the data collection.

A total of 42 people who had had a stroke took part in the study.

The overall response pattern is shown in Table 11.

Table 11. Number of experimental participants referred to study and reasons for not taking part

	Number of referrals
Number of people referred to the study	**92**
Number who refused to take part	3
Number who did not a return health screening questionnaire	8
Number ineligible	24
Had not retained licence whilst awaiting decision from DVLA	*10*
Already had a DVLA decision	*9*
Prescribed disability	*3*
Live outside study area	*2*
Number who withdrew from study	12
Due to ill health	*4*
Too far to travel	*3*
Availability problems	*2*
Too anxious to take part	*1*
Did not want to come alone	*1*
Caring for sick family member	*1*
Number who postponed their assessment date	3
Number assessed	**42**

The 42 participants were between 38 and 81 years old, with a mean age of 65.2 years (SD 9.3), and most were male (37). The mean number of years driving experience was 40.6 (SD 11.7). Most participants were more than 1 year post stroke, as illustrated in Table 12. The minimum time since stroke on the day of the assessment was 2 months and the maximum was 11 years.

Details of 26 participants' stroke were provided by their doctor. These details are summarised in Table 12. Due to the large amount of missing data, differences by stroke type were not examined in the analyses.

Table 12: Details of participants' latest stroke

	Number of participants
Time since stroke (n=41)	
1 to 6 months	5
7 to 12 months	8
13 to 24 months	9
24 to 36 months	5
More than 36 months	14
Type of stroke (n=25)	
Ischaemic	16
Haemorrhagic	7
Unknown	2
Side of lesion (n=23)	
Left	11
Right	10
Bilateral	2
Visual disturbance (n=26)	
Hemianopia	3
Quadrantanopia	1
Diplopia	2
No field defect but limitation of lateral gaze bilaterally	1
None	14
Unknown	6
Motor disturbance (n=26)	
Monoparesis	2
Hemiparesis	9
Hemiplegia	7
Right sided weakness	2

Right arm weakness	1
Facial weakness and weakness in hand	1
Broad clumsy gait	1
None	3
Unknown	3
Ataxia (n=24)	
Ataxia	7
None	10
Unknown	7
Sensory disturbance (n=25)	
Sensory loss	6
Sensory inattention	1
Not specified	1
None	9
Unknown	8
Higher Cortical disturbance (n=25)	
Dysphasia	3
Visuospatial dysfunction	2
Dyscalcilia, sequencing problems. Problems in new learning and retention.	1
Expressive	1
Poor balance	1
None	15
Unknown	5

All the experimental participants completed the SDSA and 41 completed a driving assessment. Thirty-nine experimental participants completed the UFOV.

4.4.1.2 Control participants

The control participants were recruited from the TRL database of volunteers to match the experimental participants, where possible, by age, sex and driving experience. The mean age of the control group was

63.6 years (SD 9.4) and ranged from 39 to 78 years. Thirty-eight of the 46 control participants were male and on average had 42.5 years driving experience (SD 9.9).

Of the 46 controls, 34 completed both the SDSA and the driving assessment and 12 only completed the SDSA. Thirty controls also completed the UFOV.

Table 13 shows the characteristics of both the experimental and control group.

4.4.1.3 Group differences

Differences between the experimental and control groups were investigated for each continuous variable using independent sample t-tests and the Chi-square test for the variable 'sex'. This showed that there were no statistical significant differences for the matching variables age, sex and years driven. There was a statistically significant difference between experimental and control groups for estimated miles driven per week ($p<0.05$[2]) prior to the stroke for experimental participants. The difference for estimated miles per annum was approaching significance[3]. However, mileage estimates were only available for a sub-sample and it is likely the estimates are inaccurate, especially for the experimental group who may not have driven for some time.

[2] t-test = 2.45, 44 df, p = .03
[3] t-test = 1.84, 45 df, p = .07

Table 13. Participant characteristics

	Experimental (n=42)			Control (n=46)		
	n[1]	Mean (SD)	Range	n[1]	Mean (SD)	Range
Age	41	65.2 (9.3)	38 to 81	46	63.6 (9.4)	39 to 78
Sex	42	Male	37	46	Male	38
		Female	5		Female	8
Years driving experience	31	40.6 (11.7)	10 to 57	24	42.5 (9.9)	21 to 59
Estimated miles driven per week prior to stroke*	26	257.9 (197.5)	45 to 750	20	151 (88.9)	30 to 400
Estimated miles driven per annum prior to stroke	27	13468.9 (10072.9)	2340 to 39000	20	8892 (5406.4)	2600 to 22360

* Statistical significant difference between experimental and control group, $p<0.05$
[1] Values for n are the number of participants for which the information is available

4.4.2 On-road driving assessment result

An on-road driving assessment result was available for 75 participants:

- 41 experimental participants (1 missing)
- 34 control participants (12 missing)

Table 14 indicates that stroke participants were more likely to fail than controls, with 42% of the experimental group and 10% of the control group failing. This difference between the experimental and control groups was statistically significant (p<0.01).

Table 14. On-road driving result assessment by condition

On-road driving assessment result	**Condition**		**Total**
	Experimental	Control	
Fail	17	3	**20**
Pass	24	31	**55**
Total	**41**	**34**	**75**

Chi-Square = 10.13, 1 df, p = .001

Differences between the pass and fail groups were investigated using multivariate analysis of variance, MANOVA. Age and sex were entered into the model and a significant multivariate effect[4] was found due to age[5]. Age was therefore controlled for in subsequent multivariate analyses. The characteristics of the participants in pass and fail groups are summarised in Table 15 (all participants) and Table 16 (experimental group only).

[4] MANOVA F = 4.90, p = .010
[5] Univariate F = 8.34, p = .005

Table 15. Characteristics of on-road driving assessment pass and fail groups

	Pass (n=55)			Fail (n=20)		
	n[1]	**Mean (SD)**	**Range**	**n[1]**	**Mean (SD)**	**Range**
Age*	55	62.8 (9)	39 to 76	19	69.7 (8.7)	38 to 81
Sex	55	Male	49	20	Male	16
		Female	6		Female	4
Years driving experience	43	41.3 (9.8)	20 to 59	11	43.2 (14.8)	10 to 57

* Statistical significant difference between pass and fail group, $p < 0.01$
[1] Values for n are the number of participants for which the information is available

Table 16. Characteristics of on-road driving assessment pass and fail groups (experimental group only)

	Pass (n=24)			Fail (n=17)		
	n[1]	**Mean (SD)**	**Range**	**n[1]**	**Mean (SD)**	**Range**
Age*	24	62.9 (8.2)	41 to 75	16	69.4 (9.4)	38 to 81
Sex	24	Male	23	17	Male	13
		Female	1		Female	4
Years driving experience	21	40.4 (9.5)	20 to 53	9	42.3 (16.3)	10 to 57

* Statistical significant difference between pass and fail group, $p < 0.05$
[1] Values for n are the number of participants for which the information is available

A significant multivariate effect was also found when the individual elements of the driving assessment were entered into MANOVA[6]. For each element of the on-road driving assessment the pass group performed significantly better than the fail group, as shown in Tables 17. Table 18 shows that stroke participants who passed the on-road driving assessment performed significantly better on each element except selective attention.

Table 17. Mean scores for elements of the on-road driving assessment

Measure	**Mean score (SD)**		**F**	**p**
	On-road fail n= 20	On-road pass n =55		
Decision making	3.16 (0.6)	3.89 (0.2)	51.87	***
Divided attention	2.72 (1)	3.62 (0.6)	17.12	***
Spatial ability	2.83 (1.3)	3.91 (0.3)	34.03	***
Interaction with other road users	2.77 (0.8)	3.71 (0.5)	32.79	***
Information processing	2.32 (0.9)	3.49 (0.6)	34.29	***
Selective attention	2.00 (1)	2.69 (1)	4.03	*
In-traffic score	2.87 (0.6)	3.69 (0.3)	60.69	***
Age	69.7 (8.7)	62.8 (9)	7.6	**

* = p <0.05, ** = p <0.01, *** = p <0.001

[6] MANOVA F = 9.97, df = 7, p = .000 (all participants) MANOVA F = 8.39, df = 7, p = .000 (experimental only)

Table 18. Mean scores for elements of the on-road driving assessment (experimental group only)

Measure	Mean score (SD)		F	p
	On-road fail n= 13	On-road pass n =24		
Decision making	2.99 (0.5)	3.79 (0.3)	38.17	***
Divided attention	2.47 (0.8)	3.50 (0.7)	14.19	**
Spatial ability	2.60 (1.2)	3.79 (0.4)	19.41	***
Interaction with other road users	2.50 (0.6)	3.43 (0.6)	22.6	***
Information processing	2.07 (0.8)	3.26 (0.5)	27.7	***
Selective attention	1.93 (1)	2.33 (0.9)	0.64	ns
In-traffic score	2.66 (0.4)	3.53 (0.3)	56.15	***
Age	69.4 (9.4)	62.9 (8.2)	4.43	*

ns = not significant, * = $p<0.05$, ** = $p <0.01$, *** = $p <0.001$

4.4.3 Outcome of SDSA

Table 19 shows the outcome of SDSA for the experimental and control groups. It shows stroke participants were less likely to pass the SDSA than controls. The difference between experimental and control groups was statistically significant ($p<0.001$).

Table 19. Outcome of SDSA for experimental and control groups

SDSA outcome	Condition		Total
	Experimental (post stroke)	Control	
Pass	18	42	**60**
Fail	24	4	**28**
Total	**42**	**46**	**88**

Chi-Square = 23.75, 1 df, p = .000

4.4.3.1 SDSA sub-tests

A MANOVA was used to test whether there were any differences between experimental and control groups in the individual SDSA components. This indicated that there was a statistically significant multivariate effect ($p < 0.001$[7]). Therefore the univariate effects were explored and these are summarised in Table 20.

Table 20 indicates that the control group performed better than the experimental participants on all aspects of the SDSA. The differences for all except one score (Dot cancellation: False Positives) were statistically significant at $p<0.05$.

Table 20. Mean SDSA sub-test scores by condition

Measure	Mean score (SD)		F	p
	Experimental n=46	Control n=42		
Dot Cancellation: Time	526.43 (156.87)	403.30 (79.02)	22.19	***
Dot Cancellation: Errors	17.81 (24.05)	9.72 (9.87)	4.40	*
Dot Cancellation: False Positives	0.43 (1.06)	0.15 (0.60)	2.32	ns
Square Matrix: Directions Score	24.62 (10.84)	30.57 (4.00)	12.06	**
Square Matrix: Compass Score	16.79 (8.40)	26.78 (6.76)	38.16	***
Road Sign Recognition Score	6.69 (2.96)	9.00 (2.81)	14.09	***
Equation(pass – fail equation)	-0.17 (2.78)	3.55 (2.4)	45.41	***

ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

A statistically significant multivariate effect[8] was also found between those who passed and failed the on-road driving assessment. The pass group scored statistically significantly better on each SDSA sub-test except the Dot Cancellation: False Positives score, as shown in Table 21. The findings differ when the experimental group is considered alone. Experimental participants who passed the on-road driving assessment performed significantly better than experimental participants who failed

[7] MANOVA F=8.83, df = 6, p = .000

[8] MANOVA F=5.91, df = 7, p = .000 (all participants); MANOVA F=4.34, df = 7, p = .001 (experimental only)

on the Dot cancellation: time, Dot cancellation: False Positives and Road Sign Recognition test. There were no significant differences between pass and fail experimental group scores for the Dot cancellation: Errors or Square Matrices tests.

Table 21. Mean SDSA sub-test scores by on-road driving assessment result

Measure	Mean score (SD)		F	p
	On-road fail n=20	On-road pass n=55		
Dot Cancellation: Time	582.4 (191.2)	430.7 (90.1)	18.73	***
Dot Cancellation: Errors	24.2 (31.3)	10 (9.8)	8.23	**
Dot Cancellation: False Positives	0.8 (1.4)	0.2 (0.6)	6.05	ns
Square Matrix: Directions Score	23.1 (13)	28.8 (6.5)	4.5	*
Square Matrix: Compass Score	17.4 (9)	23.2 (8.8)	5.21	*
Road Sign Recognition Score	5.7 (3.4)	8.4 (2.8)	11.46	**
Equation (Pass – fail equation)	-0.61 (3.2)	2.4 (2.9)	12.59	**
Age	69.7 (8.7)	62.8 (9)	.34	**

ns = not significant, * = p <0.05, ** = p <0.01, *** = p <0.001

Table 22. Mean SDSA sub-test scores by on-road driving assessment result (experimental group only)

Measure	Mean score (SD)		F	p
	On-road fail n=17	On-road pass n=24		
Dot Cancellation: Time	612.0 (191.9)	465.2 (94.5)	8.94	**
Dot Cancellation: Errors	26.5 (33.2)	12.2 (12.9)	3.29	ns
Dot Cancellation: False Positives	0.9 (1.5)	0.1 (0.3)	6.08	*
Square Matrix: Directions Score	21.5 (13.6)	26.5 (8.3)	1.34	ns
Square Matrix: Compass Score	16.1 (8.9)	17.0 (8.3)	0.04	ns
Road Sign Recognition Score	5.5 (3.0)	7.4 (2.6)	4.22	*
Equation (Pass – fail equation)	-1.1 (3.1)	0.4 (2.4)	2.28	ns
Age	69.4 (9.4)	62.9 (8.2)	5.36	*

ns = not significant, * = p <0.05, ** = p <0.01, *** = p <0.001

4.4.4 UFOV risk categories

The following tables present the UFOV risk categories[9] by condition (Table 23), SDSA outcome (Tables 24 and 25) and on-road driving assessment outcome (Tables 26 and 27). Table 23 shows that the experimental group was more likely to be in a higher UFOV risk category than the control group and the difference between groups was statistically significant (p<0.01)[10]. Tables 24 and 25 indicate that participants who failed the SDSA were more likely to be in higher UFOV risk categories. These differences between pass and fail groups were statistically significant (p<0.001 and p<0.05). Table 26 shows that when control and experimental participants are considered together, those who failed the on-road driving assessment were more likely to be in higher

[9] The risk categories automatically computed by the pc version 5.6 of UFOV were found to incorrectly group some participants. The categories were recalculated using information provided by Visual Awareness, Inc., who produce and licence the UFOV, and were used in the analysis.
[10] Chi-square = 16.17, df = 4, p = .003

UFOV risk categories. Note, however, that nine of the 17 people assigned to the UFOV high or very high risk category passed the on-road test. For the experimental group only there was no significant difference in UFOV risk category between SDSA pass and fail groups, see Table 27.

Table 23. UFOV risk category by condition

UFOV risk category	**Condition**		**Total**
	Experimental (n=39)	Control (n=32)	
Very low risk	10	21	**31**
Low risk	9	8	**17**
Low to moderate risk	4	1	**5**
Moderate to high risk	1	0	**1**
High risk and very high risk	15	2	**17**

Chi-square = 16.17, df = 4, p = .003

Table 24. UFOV risk category by SDSA outcome

UFOV risk category	**SDSA outcome**		**Total**
	Fail (n= 24)	Pass (n= 47)	
Very low risk	4	27	**31**
Low risk	4	13	**17**
Low to moderate risk	2	3	**5**
Moderate to high risk	1	0	**1**
High risk and very high risk	13	4	**17**

Chi-square = 22.73, df = 4, p = .000

Table 25. UFOV risk category by SDSA outcome (experimental group only)

UFOV risk category	SDSA outcome		Total
	Fail (n= 22)	Pass (n= 17)	
Very low risk	4	6	**10**
Low risk	3	6	**9**
Low to moderate risk	1	3	**4**
Moderate to high risk	1	0	**1**
High risk and very high risk	13	2	**15**

Chi-square = 11.0, df = 4, p = .026

Table 26. UFOV risk category by on-road driving assessment result

UFOV risk category	On-road driving assessment result		Total
	Fail (n= 18)	Pass (n= 50)	
Very low risk	4	24	**28**
Low risk	2	15	**17**
Low to moderate risk	3	2	**5**
Moderate to high risk	1	0	**1**
High risk and very high risk	8	9	**17**

Chi-square = 13.39, df = 4, p = .010

Table 27. UFOV risk category by on-road driving assessment result (experimental group only)

UFOV risk category	SDSA outcome		Total
	Fail (n= 15)	Pass (n= 23)	
Very low risk	2	7	**9**
Low risk	1	8	**9**
Low to moderate risk	3	1	**4**
Moderate to high risk	1	0	**1**
High risk and very high risk	8	7	**15**

Chi-square = 9.0, df = 4, p =.061

4.4.5 Predictive validity

This project assessed the predictive validity of the SDSA as a screening tool that would identify and fail people who would otherwise fail an on-road driving assessment. In doing so the sensitivity, specificity and predictive values of the test were examined.

Sensitivity – the probability that a person who fails the road test would fail the SDSA. A test with high sensitivity has few 'false negatives' – that is, few people who fail the road test would pass the SDSA. High sensitivity is desirable, since we wish to screen out most of the people who would fail the on-road driving assessment, thus reducing the costs (monetary and other) associated with failures of the road test. Low sensitivity would mean that the SDSA would pass a substantial proportion of the people who fail the road test. The consequence would be that substantial proportions of road tests would be failed, and their associated costs experienced. However, on the assumption that the road test stands as a fully valid predictor of safety[11] there would be no direct adverse consequences for safety since the on-road test would still be there to identify people who are not safe to drive. Thus, high sensitivity is desirable, but there are no severe consequences for accepting a test with lower sensitivity.

Specificity – the probability that a person who would pass the road test would also pass the SDSA. Low specificity would mean that the SDSA would fail a substantial proportion of the people who would have gone on to pass the road test. This would be ethically unacceptable, since it would mean that such people would be delayed or prevented from returning to driving.

Positive predictive value (PPV) – the probability of a person failing the SDSA also failing the road test. The PPV captures the ethical dimension referred to above, since the value of one minus PPV gives the proportion of those who fail the SDSA who would otherwise go on to take and pass the road test. As with specificity, the SDSA needs a high value for PPV if it is to be ethically acceptable as a screening test operating as gatekeeper to the on-road test.

Negative predictive value (NPV) – the probability of SDSA passers also passing the road test.

Table 28 shows the outcomes of the on-road driving assessments and the SDSA for the experimental group in this study. Twelve of the 17

[11] This assumption is hardly likely to be fully valid, but it does reflect the model underlying this study – i.e. that the road test provides the decision on returning to drive, and the SDSA is being evaluated as a gatekeeper to the road test. Alternative models are discussed in section 5.

people who failed the on-road test were identified (failed) by the SDSA, giving a sensitivity of 71%). Twelve of the 24 drivers who passed the on-road test also passed the SDSA giving a specificity of 50%. Twenty-four of the 41 participants were correctly classified by the SDSA – an 'accuracy' of 59%. These results are presented in Table 29.

Table 28. SDSA outcome by on-road driving assessment result for the experimental group

SDSA outcome		On-road driving assessment result		Total
		Fail	Pass	
		+	-	
Fail	+	12	12	24
Pass	-	5	12	17
Total		17	24	41

Table 29. Test performance indices for the SDSA (experimental group only)

	Percentage (95% Confidence Intervals[12])
Sensitivity	71
Specificity	50
Accuracy	58
Positive Predictive Value	50 (30 to 70)
Negative Predictive Value	71 (45 to 88)

Twelve of the 24 people who failed the SDSA also failed the on-road test, giving a PPV of 50%, the 95% confidence interval for this value being 30% to 70%. This means that it is very unlikely that more than 70% of those who failed the SDSA would fail the road test. We can, therefore, be almost certain that if the SDSA were to be used in its present form as a gatekeeper for the on-road test, at least 30% of the people it failed would be incorrectly failed – in that they would be denied (or delayed) access to an on-road test that they would have passed. This assumes that the sample of post-stroke drivers in this study is representative of

[12] 95% Confidence Intervals calculated using Wilson's method and adjusted for small sample sizes.

the drivers who would come forward to be assessed by the SDSA if it were adopted. The SDSA would correctly identify 12 in 17 of the people who would fail the road test. This means that if introduced, it could reduce the number of failed road tests by 71%.

The outcomes of the on-road driving assessments and the SDSA for the control group are presented in Table 30. It shows that few control participants failed the on-road driving assessment or the SDSA. Of the three control participants who failed the on-road driving assessment one also failed the SDSA, whilst another participant who failed the SDSA passed the on-road driving assessment.

There is no reason to expect SDSA outcome to predict on-road outcome for the control sample: the discriminant function analysis on which the scoring system is based was derived for a sample of stroke patients referred for an assessment of their fitness to drive. However, it would be unfortunate if a large proportion of healthy controls did fail the SDSA, especially if this turned out to predict their failure of the on-road assessment. In such circumstances it might be argued that applying the SDSA to stroke patients would be unfair, as it would place a hurdle in the way of their driving that healthy drivers do not have to clear. The findings in Table 30 indicate that this problem does not occur in practice.

Table 30. SDSA outcome by on-road driving assessment result for the control group

SDSA outcome		On-road driving assessment result		Total
		Fail	Pass	
		+	-	
Fail	+	1	1	2
Pass	-	2	30	32
Total		3	31	34

4.4.5.1 SDSA ROC curve

The analyses above take the pass-fail outcome of the SDSA as determined by the scoring system described by Nouri and Lincoln (1992). It appears that using this scoring system, the SDSA could usefully reduce the number of failed road tests by up to 70%, but that an unacceptably[13] large number of people would incorrectly have their

[13] This study cannot determine what an acceptable figure for PPV would be, since this is a matter for judgement rather than science. However, we assume here that

access to a Mobility Centre assessment denied or at least delayed. The question arises of whether the pass-fail criterion could be changed in a way that improves the PPV and specificity of the SDSA without sacrificing too much sensitivity. Devising alternatives to the pass and fail equations is outside the scope of this study but it was decided to investigate the effect of changing the pass-fail cut-off score. Currently the cut-off score is zero. That is, if the difference between the pass equation and fail equation scores is positive, a pass is awarded, and if it is negative a fail is awarded. However it is possible to vary this and use other cut-off scores for this variable.

One way to show the results of such an investigation is to plot the Receiver Operating Characteristic (ROC) curve for the test. This plots sensitivity against one minus specificity (and each point on the curve corresponds to a different pass-fail cut-off point being applied to the test). Figure 4 shows the ROC curve derived for the SDSA for the experimental group. Sensitivity and 1-specificity are shown as percentages. For example, when the cut-off point is zero (0.0), sensitivity is 0.71 or 71% and 1-specificity is 1 - 0.50 = 0.50 or 50%. The sensitivity and 1-specificity was the same for some cut-off points, if there were the same number of participants passing or failing the SDSA at that cut-off point e.g. -1.0 and -1.5.

30% SDSA failers being incorrectly failed (in the sense that they would pass the on-road test) is unacceptable; and this figure is at the extreme end of the confidence interval, so the true proportion of incorrect failures is probably higher, perhaps very much higher, than 30%. Of course, a relatively high rate of incorrect SDSA failures (i.e. a relatively low PPV) becomes more acceptable if the penalty for being incorrectly failed becomes less severe. While it is probably unacceptable for 30% of SDSA failers to be incorrectly and permanently denied access to the road test, it may be acceptable to delay their access if introducing the test brings other benefits. This point is taken up in section 5.

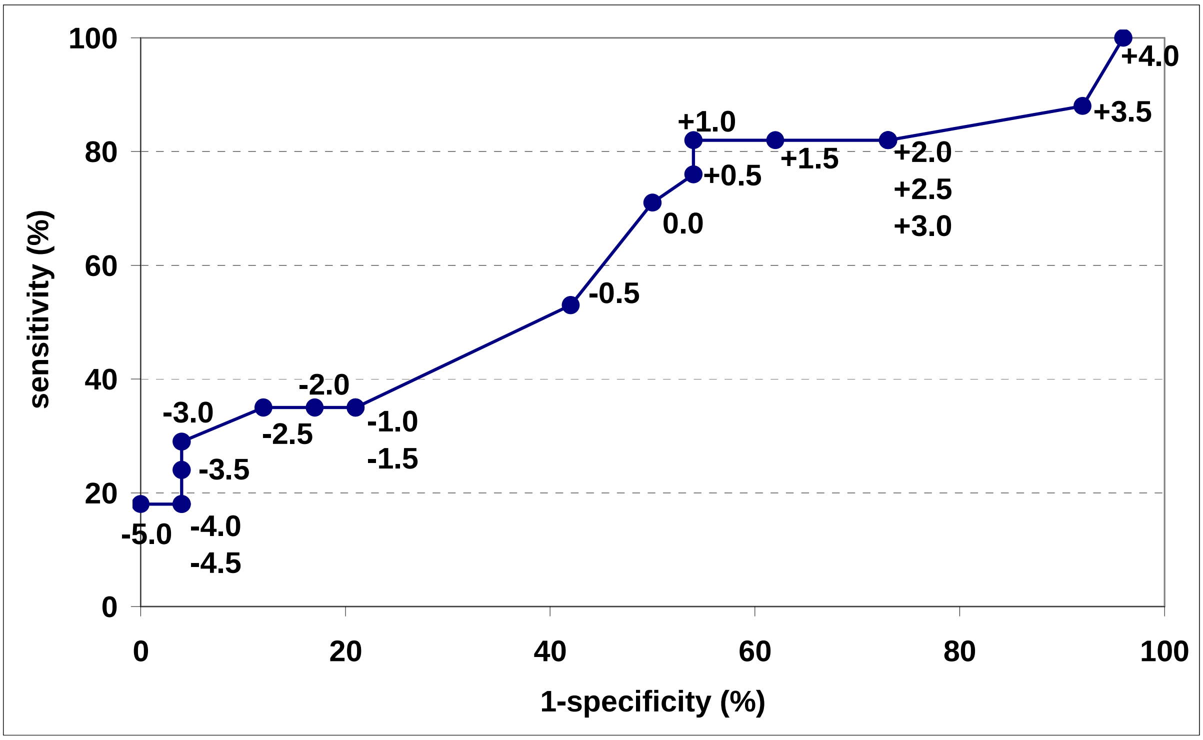


Figure 4. SDSA ROC Curve (experimental group)

As the test is made more difficult to pass, sensitivity increases and specificity reduces. As the test is made easier to pass, the opposite happens. For example, by adopting -3 instead of zero as the pass-fail cut-off, specificity increases to 96% and sensitivity reduces to 29%. At this point, PPV would be 83% (95% CI = 34% to 98%). With this cut-off value it is therefore possible that the SDSA would incorrectly fail acceptably few people, but still be able to provide a worthwhile reduction in road test failures by correctly identifying 29% of the people who would otherwise fail the road test. The wide confidence interval for PPV means, however, that we are still very uncertain of the true value of the incorrect failure rate (1-PPV). Table 31 summarises the sensitivity, specificity and PPV for a number of cut-off points.

Table 31. Test performance indices for various cut-off points of the difference between pass and fail equations score variable (experimental group only)

Cut-off point	Percentage		
	Sensitivity	Specificity	PPV (95% CI)[14]
0	71	50	50 (30 to 70)
-0.5	53	58	47 (26 to 70)
-1	35	79	55 (25 to 81)
-2	35	83	60 (28 to 85)
-3	29	96	83 (34 to 98)

4.4.5.2 NorSDSA predictive equations

An alternative way to score the SDSA is provided by the NorSDSA predictive equations (Lundberg et al. 2003).

Table 32 shows the outcomes of the on-road driving assessments and the SDSA for the experimental group in this study, calculated using the NorSDSA predictive equations. Three of the 17 people who failed the on-road test were correctly identified using the NorSDSA predictive equations, giving a sensitivity of 18%. Only one of the 24 drivers who passed the on-road driving assessment failed the SDSA using the NorSDSA equations giving a specificity of 98%. Twenty-six of the 41 participants (66%) were correctly classified by the SDSA using the NorSDSA predictive equations.

Three of the four people who failed the SDSA, using the NorSDSA equations, also failed the on-road test, giving a PPV of 75%, the 95% confidence interval for this value being 16% to 98%.

Table 33 summarises the indices of test performance for the experimental group. As with the results obtained above for the standard SDSA it can be seen that useful sensitivity combined with acceptably high PPV appears possible here, but the confidence interval for PPV is again very wide.

[14] 95% Confidence Intervals calculated using Wilson's method and adjusted for small sample sizes.

Table 32. NorSDSA outcome by on-road driving assessment result for the experimental group

NorSDSA outcome		On-road driving assessment result		Total
		Fail	Pass	
		+	-	
Fail	+	3	1	4
Pass	-	14	23	37
Total		17	24	41

Table 33. Test performance indices for the SDSA calculated using the NorSDSA predictive equations for the experimental group

	Percentage (95% Confidence Intervals[15])
Sensitivity	18
Specificity	96
Accuracy	66
Positive Predictive Value	75 (16 to 98)
Negative Predictive Value	62 (46 to 76)

4.4.6 Concurrent validity

To consider concurrent validity of the SDSA, the correlation between the SDSA and UFOV was investigated. As mentioned in section 3.5.5 the UFOV has been shown by its developers to be predictive of crash rate in older drivers and so can distinguish between at least some types of safe and unsafe drivers. It has not, however, been used to assess drivers following a stroke. It can therefore provide only a very imperfect indication of concurrent validity.

In addition the correlation between the on-road driving assessment and the SDSA and UFOV measures was also examined to investigate if and how the different elements were associated. The correlation matrix is presented in Table 34 with values for the experimental group shown below the diagonal and values for the whole sample shown above it.

[15] 95% Confidence Intervals calculated using Wilson's method and adjusted for small sample sizes.

Experimental group

For the experimental group, Table 29 shows that the difference between the SDSA pass – fail equations (SDSA outcome score) was significantly correlated with the UFOV risk category (r = -0.54, p<0.001). Three components of the SDSA were significantly correlated with the UFOV risk category at the 5% level: the square matrix directions test[16]; square matrix compass test[17] and the road sign recognition test[18].

Several SDSA components were significantly correlated with UFOV subtests, the highest value (r = 0.68, p<0.001) being for UFOV subtest 1 (processing speed) and the SDSA Dot Cancellation Time. SDSA Square Matrix Directions test showed significant correlations with all three UFOV subtests, but SDSA Dot Errors score was not significantly correlated with any of the UFOV subtests. SDSA Dot Errors, SDSA Dot Time and SDSA Dot False Positives were not significantly correlated with UFOV overall risk category.

Whole sample

For the whole sample a generally similar pattern is seen, with values of correlation coefficients tending to be higher than in the experimental sample (as would be expected from the generally wider variable ranges likely to occur in the whole sample). Some additional correlations reach statistical significance, so that all the SDSA components except Dot cancellation errors become significantly correlated with UFOV risk category.

The time recorded to complete the dot matrix test[19] and number of false positives[20] were only significantly correlated with the first UFOV sub-test.

The pattern of correlations provides moderate support for the concurrent validity of the SDSA – if the UFOV is itself accepted as valid for assessing whether a person is ready to return to driving following a stroke.

It is also instructive to examine the correlations between UFOV scores and driving assessment scores.

For the experimental sample there was no statistically significant correlation between UFOV risk category and overall driving assessment score. However, UFOV subtask 1 (processing speed) was significantly correlated with driving assessment score[21]. There were also significant correlations between individual UFOV and individual driving assessment

[16] Correlation r = –0.49, p<0.01
[17] Correlation r = –0.51, p<0.01
[18] Correlation r = –0.37, p<0.05
[19] Correlation r = 0.68, p<0.001
[20] Correlation r = 0.47, p<0.05
[21] Correlation r = –0.47, p<0.01

subtests. However, there was no significant correlation between the UFOV subtest1 and item 2 of the driving assessment, both of which were held to measure divided attention. Likewise, the UFOV and driving assessment items held to measure selective attention showed no statistically significant correlations.

With the whole sample (experimental and control) all the correlations between UFOV and the driving assessment became statistically significant.

The above pattern of results suggests that although the performance characteristics tapped by the UFOV and driving assessment show a fair degree of overlap there may be particular characteristics of post stroke individuals that differentially affect the two tests, such that for these people the correlation between the two sets of results are weakened.

Table 34. CORRELATION MATRIX: Driving Assessment Measures, SDSA Measures, UFOV measures

Variable[1]	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.
1. DDecision	-	.59***	.72***	.78***	.72***	.30*	.92***	-.57***	-.44***	-.37**	.42***	.43***	.36***	.51***	-.37**	-.61***	-.39**	-.37**
2. DDivAtt	.54***	-	.59***	.54***	.50***	.27*	.63***	-.48***	-.23	-.11	.38**	.31**	.27*	.40***	-.38**	-.37**	-.34**	-.32**
3. DSpatial	.65***	.56***	-	.59***	.56***	.26*	.72***	-.30**	-.10	-.04	.14	.19	.14	.24*	-.36**	-.39**	-.34**	-.34**
4. DInteract	.69***	.48**	.49**	-	.72***	.20	.87***	-.32**	-.14	-.16	.23	.43***	.29*	.45***	-.39**	-.33**	-.35**	-.38**
5. DInfoPro	.68***	.47**	.53***	.68***	-	.30*	.89***	-.44***	-.27*	-.34**	.34**	.33**	.36**	.43***	-.42**	-.45***	-.37**	-.41**
6. DSelAtt	.17	.19	.19	-.01	.18	-	.38***	-.43***	-.26*	-.18	.30**	.29*	.25*	.36*	-.32**	-.32**	-.35**	-.36**
7. DRVScore	.91***	.60***	.68***	.83***	.88***	.21	-	-.50***	-.34**	-.35**	.34**	.42***	.34**	.49***	-.43***	-.46***	-.41**	-.45***
8. SDot-Time	-.50***	-.46**	-.17	-.10	-.39*	-.32*	-.38*	-	.30**	.48***	-.58***	-.42***	-.43***	-.63***	.38**	.62***	.36**	.37**
9. SDot-Errors	-.39*	-.19	.01	.03	-.21	-.29	-.27	.37*	-	.53***	-.41***	-.31**	-.22*	-.31**	.15	.25*	.23	.21
10. SDot-False	-.42**	-.13	.02	-.12	-.49**	-.18	-.45**	.55***	.63***	-	-.38***	-.29**	-.27*	-.34**	-.28*	.41***	.30*	.28*
11. SSqDirection	.32*	.31	-.01	-.01	.27	.33*	.20	-.65***	-.36*	-.45**	-	.61***	.49***	.67***	-.54***	-.59***	-.53***	-.46***
12. SSqCompass	.18	.20	-.09	.09	.11	.08	.13	-.31*	-.23	-.33*	.60***	-	.51***	.91***	-.61***	-.33**	-.60***	-.60***
13. SRdSignRec	.29	.25	.01	.10	.30	.13	.22	-.52***	-.19	-.38*	.46**	.31*	-	.77***	.47***	-.43***	-.44***	-.47***
14. SDSA Outcome score	.35*	.35*	-.00	.12	.27	.18	.25	-.65***	-.29	-.46**	.73***	.86***	.71***	-	-.63***	-.47***	-.60***	-.62***
15. UFOV risk category	-.21	-.34	-.25	-.22	-.36*	-.27	-.29	.26	.05	.23	-.49**	-.51**	-.37*	-.54***	-	.38**	.93***	.89***
16. Usubtest1	-.61***	-.34*	-.34*	-.29	-.51**	-.38*	-.47**	.68***	.22	.47*	-.57***	-.31	-.52**	-.55***	.34*	-	.34**	.31*
17. Usubtest2	-.23	-.31	-.21	-.13	-.27	-.23	-.23	.20	.14	.24	-.50**	-.49**	-.31	-.48**	.90***	.33*	-	.83***
18. Usubtest3	-.22	-.36*	-.23	-.19	-.37*	-.29	-.30	.19	.14	.26	-.44**	-.51**	-.38*	-.52**	.94***	.33*	.86***	-

Note: Correlation coefficients above the diagonal relate to the whole sample (experimental and control participants) and correlation coefficients below the diagonal relate to the experimental group only.

[1] A list of variable names can be found in Appendix A. The shaded area indicates the SDSA measures.

Statistical significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.4.7 Face validity

A finding from the questionnaire survey that was confirmed by the telephone interviews and unprompted comments made by participants in the validity study were that clients do not readily accept the relevance of the SDSA. This seems to be due to the abstract nature of the SDSA. Clients do not see the relevance of ‘pen and paper’ tests to driving.

SDSA users interviewed as part of the survey explained that, because the SDSA was specifically linked to driving, patients expected more driving orientated tasks. Clients more readily accepted the results of other cognitive tests because they were explained to the client as simply testing the functioning of a certain area of the brain rather than an entire skill such as driving.

Some SDSA users have found that if it is explained properly in terms of the skills it is testing and how these relate to driving then clients are more inclined to accept the result of their assessment. However those that fail the SDSA are sometimes difficult to convince of its relevance when they themselves are certain they are fit to return to driving.

SDSA users generally believed the road sign recognition test had greater face validity than the other component tests. Several SDSA users commented on the unrealistic nature of the roundabout cards in the square matrix compass test as it is shown to have 8 exits. Although this may be unrealistic it would appear necessary for the purpose of this test and the format and working of a roundabout should be familiar to the clients regardless of the number of exits.

5 Discussion and conclusions

This project consisted of four studies to investigate the current use of the SDSA in the UK and examine its validity and reliability as a screening tool to identify drivers who should not return to driving at that time.

The survey of users suggests that the SDSA is not widely used. This is consistent with, previous studies of occupational therapy practice, which have highlighted that few therapists use standardised published assessments (Walker et al 2003, Sackley and Lincoln, 1996). Those healthcare professionals who are currently using the SDSA do so infrequently and not routinely with all patients. There appears to be little or no consensus among users on the circumstances under which the SDSA should be used and this is an issue that needs to be addressed. In principle (and if its performance were satisfactory), the screening tool could be used to assess whether a driver should be referred to a mobility centre for a full driving assessment or it could be used by the DVLA to determine fitness to drive. It could also be used in combination with the on-road driving assessment to determine fitness to drive.

Most of the current users who responded to the survey were occupational therapists, who felt the SDSA is ‘useful’ in helping to assess a client's fitness to drive but that on its own cannot be used to predict driving performance. It was, however, regarded as useful to help identify those clients who need to undergo a full driving assessment. The inter-rater reliability study found that rather than screening clients for further assessment, all the 28 SDSA users recommended that all three clients shown in the video, even those who failed the SDSA, should undergo further assessment, whether it be physical, cognitive or a driving assessment. The raters did not envisage using the outcome of the SDSA alone to advise clients that they should not return to driving at this time, and did not appear confident in its validity.

When the ease of use of each test was considered, the SDSA users in the survey reported that the square matrices test was the most difficult, partly because of shading of the compass points; and the dot matrix was the easiest to administer and interpret, although an overlay is needed to score it accurately.

One issue raised by the SDSA users that could impact on the predictive ability of the test was that some SDSA users provided additional instructions and information to clients who did not understand what it was they were supposed to do. This undermines the SDSA’s status as a standardised test and may reflect the lack of training in psychometric testing given to occupational therapists. The issue has been partially addressed in the revised SDSA manual (Lincoln et al., 2004). The revised manual includes phrases that should be used if the client does not understand the task, instructs the user to repeat the instructions only once and emphasises not to elaborate the instructions. Providing training in how to administer the SDSA should also help overcome these difficulties.

Another issue addressed in the revised manual is the presentation of the cards in the road sign recognition test. Some of the users in the inter-rater reliability study commented that they placed the road sign cards in a pile in front of the client, rather than spread across the table in front of the client as shown in the video. If cards are spread across the table, clients may find it easier to select the matching sign for the road situation cards by looking at the situation card first and then the signs. The revised manual states the road signs should be spread below the situations. When the test was administered in the validity study the cards were placed in a pile rather than spread across the table in this way. The validity study procedure was not amended following the publication of the revised manual since it was judged preferable to be consistent in all the tests.

The scoring of the SDSA had high inter-rater reliability with an Intra-class Correlation Coefficient (ICC) of at least 0.8 for each the components of the SDSA. The accuracy of the scoring in the inter-rater reliability study showed scope for improvement: 11 of the 84 SDSA score sheets contained calculation errors made by the raters. Most of these were calculation errors when multiplying or summing the SDSA components in the equations. Other errors included recording the time for the dot cancellation test in minutes rather than seconds and transcribing an incorrect figure into an equation from the individual scores. In the majority of cases the errors did not affect the overall result (i.e. whether the pass or fail equation score was greater). The provision of an electronic version of the scoring sheet may improve the accuracy of scoring.

The survey of SDSA users suggested that the accuracy of the SDSA scoring procedure is compromised by not providing scoring criteria for the ‘road sign recognition’ test. The revised manual now illustrates the answers to this test so SDSA users do not have to use their own knowledge to score the clients’ responses. Few inconsistencies in scoring were highlighted by users; one rater in the inter-rater reliability study included the practice line in the dot matrix score but the manual states that this should not be included. The scoring of all correctly placed vehicles in a pile of cards in the compass test could be clarified in the manual. It is theoretically possible in such a case to achieve the maximum score and not correctly place all the cards because there are more cards than spaces. The importance of shuffling the cards before each test could also be reinforced.

The interpretation of the results was less consistent than scoring and was found to vary between raters in the inter-rater reliability study. A mixed outcome was found for two clients where the difference between the pass and fail equations was small with some raters either passing or failing the client and others using a ‘borderline’ category. The SDSA user manuals (Nouri and Lincoln, undated; Lincoln et al., 2004) do not define a ‘borderline’ category, although the revised manual quotes a borderline category suggested by a study of the Nordic version of the SDSA of -0.49 to 0.49 for the difference between pass and fail equations. (Lundberg et al., 2003). Unlike previous validity studies (Nouri et al., 1992; Lundberg et al., 2003) the driving assessment in this validity study was based on part of the MAVIS driving assessment and did not use a borderline result category so marginal scores could not be investigated in the same way. The SDSA outcome score for the two clients assigned a ‘borderline’ result by the raters in the inter-rater reliability study ranged from -1.11 to +1.03. The range of SDSA outcome score for those who failed the SDSA but passed the driving assessment was broad (-4.97 to -0.14) so is not useful in establishing a borderline category. The ROC curve indicates that a borderline category covering outcome scores between -1.0 and +1.0 would improve the accuracy of prediction. If a borderline category is to be established it needs to be clearly defined for users and indicate the advice that should be given to the client.

To summarise, the findings of the survey of users and inter-rater reliability study indicate that inter-rater reliability of scoring is high but there are differences in the administration of the SDSA and with the interpretation of its results which are cause for concern.

The test-retest study carried out by Lincoln and Fanthome (1994) used a sample of 36 patients and found that whilst there were significant practice effects for the individual test components none of the participants changed from a fail to a pass on second administration. However, the sample did not consist of people who had been referred for driving assessment so was not representative of the population the test was developed for.

Although it is difficult to extrapolate the results obtained on such a small sample (n=6), the results of our test-retest study show reasonable levels of agreement between scores on first and second administration for the dot cancellation time, dot cancellation test errors and square matrix test. A lower level of agreement was found for the square matrix compass test and road sign recognition test reflecting changes in the rank order of participants at retest. Changes in mean scores show that performance on several components of the test and overall outcome score (pass – fail equation score) significantly improve either with task familiarity (on second administration in our case) or with minimal training. This susceptibility to performance improvement was most importantly exemplified in one of the two cases where SDSA was failed on first administration, when the volunteer passed on second administration. This suggests test familiarity may be enough to change a negative outcome into a positive one, though, of course, random variation (unreliability) in test performance will play its part here. The participant whose outcome changed between first and second administration had outcome scores of -0.69 and +4.16 on first and second administrations respectively. The first score was closer to a pass than that of the participant who failed on both occasions. The participant who changed

SDSA outcomes was 8 months post-stroke at the time of first administration, and it is possible that the change in SDSA outcome might reflect an improvement in this person's condition. The other five participants were between 11 months and 9 years.

The pass-fail outcome of the second and fifth (post-training) administration of the SDSA was identical for all volunteers. At least in this small sample, and despite significant or near-significant improvements in several test components, the improvements did not amount to a change of pass/fail outcome.

It should be noted that a serious limitation of our test-retest study was the very small sample. All of the participants in the study had passed the driving assessment on the day of first administration of the SDSA suggesting the sample was biased to those more able. This means that the results cannot be generalised. The results may have been different if the study had been carried out with a representative sample of the experimental group.

A potential problem with introducing screening tools for a subset of the driving population is that they might, if applied to the general population of drivers, screen out a substantial proportion of those drivers. However, the study reported here found that very few of the control sample failed either the SDSA, the on road assessment, or the UFOV.

To examine predictive validity of the SDSA for the experimental group, the sensitivity, specificity and predictive values of the test were examined. The control group was excluded from these analyses which needed to be based on a population for which the test was designed (stroke patients). In particular, the predictive values are affected by prevalence, and the inclusion of the control group affects this.

The results indicated that sensitivity of the SDSA was reasonably high at 71% but specificity was low at 50%. The Positive Predictive Value (PPV) for the SDSA was also low at 50% (95% CI 30 to 70%). These figures are lower than those reported in other studies by Nouri and Lincoln (1993) (75% sensitivity and 89% specificity), Brown (2000) (80% sensitivity and 73% specificity), or Lundberg et al., (2003) which found sensitivity of 70% and specificity of 67%.

It should be noted that Nouri and Lincoln (1993) examined the predictive ability of the test to detect *passes* rather than *fails*. In this study we were asked to examine the test as a screening tool which would act as a gatekeeper to the on-road driving assessment. Clients who fail the SDSA would be advised not to drive and those who pass would be referred for further (on-road) assessment to determine if they were safe to drive. Therefore, it is important the test should correctly predict those who would fail an on-road driving assessment as they would be denied or delayed access to an on-road driving assessment. However, even when the sensitivity and specificity for the Nouri and Lincoln (1993) sample is recalculated for detecting fails (sensitivity = 84%, specificity = 75% and PPV = 89%) the indices are still higher than those found in the present study.

The differences between the results of this study and previous studies may be due to sampling differences: the mean time since stroke is longer in the present study than in other studies. It may also be related to age. The experimental sample in this study was older than the Nouri and Lincoln sample. The experimental sample also included three participants with Hemianopia whilst Nouri and Lincoln (1992) excluded participants with these conditions from their study. The differences may also be because the driving assessments were scored differently and the driving performance criteria may also have differed: it is not clear what aspects of performance were evaluated in Nouri and Lincoln, (1993) or Lundberg et al., (2003).

Generally, if a test is to be used to regulate access to an on-road driving assessment or be used for licensing purposes, the PPV of the test should be very high. The upper 95% confidence limit for the PPV in this study was 70% which means that we can be almost certain that if the SDSA were used in its present form more than 30% of people who fail it would actually pass the on-road driving assessment. The levels that should be considered acceptable are open to debate and will depend on the consequences to the individual of 'incorrectly' failing the SDSA. However, incorrectly denying 30% or more of those who fail the SDSA access to an on-road driving assessment seems unlikely to be acceptable.

Given the low PPV for the SDSA using the existing predictive equations with a cut-off point of zero for difference between pass and fail equations, the sensitivity, specificity and PPV of various other cut-off points were explored. This highlighted that if the SDSA was made easier to pass by using a cut-off point of –3, the number of False positives reduces (the number of participants who failed the SDSA and passed the on-road driving assessment) and the PPV increases to 83% (95% confidence interval 34% to 98%). The sensitivity of the test reduces to 29% and specificity increases to 96%. With this cut off value it is therefore possible that acceptably few people would be incorrectly failed by the SDSA, which would still produce a worthwhile reduction in road test failures by correctly identifying 29% of the people who would fail the road test. The wide confidence interval for PPV means however, that we are still very uncertain of the rate of incorrectly failing the SDSA.

The consequences of using NorSDSA predictive equations are similar. When using the NorSDSA predictive equations to calculate the SDSA, the test has a sensitivity of 18% and specificity of 96%. The PPV increases to 75% (95% confidence interval 16 to 98%). The upper confidence limit for this value suggests that it is possible that few clients would incorrectly fail the SDSA and be denied access to a driving assessment, though the central estimate of PPV (75%) is more likely, and less favourable. A high proportion of those who pass the SDSA would go on to fail an on-road driving assessment, but 18% of road test failers would be correctly identified and advised not to take the road test.

The UFOV was used in the study to provide an indication of concurrent validity. The test has been shown by its developers to be predictive of crash rate in older drivers and so can distinguish between at least some types of safe and unsafe drivers. The pattern of correlations found in the present study provides moderate support for the concurrent validity of the SDSA – if the UFOV is itself accepted as valid for assessing whether a person is ready to return to driving following a stroke. However, it should be noted that such an assumption may not be well-founded. First, the UFOV's ability to predict crash-rates of post-stroke drivers is not known, secondly (and at the risk of some circularity of argument) the pattern of correlations found in this study between the UFOV and on-road driving assessment scores suggests that although the performance characteristics tapped by the UFOV and driving assessment show a fair degree of overlap, there may be particular characteristics of post-stroke individuals that differentially affect the two tests, such that for these people the correlations between the two sets of results are weakened.

Face validity of the SDSA was highlighted as an issue in the survey of SDSA users and by participants in the validity study. It appears many clients do not understand how a pen and paper based test can assess their driving ability. It appears the criticism relates to all tests that do not directly observe the driver; not just those used in the SDSA. It is likely that if the SDSA were used for licensing purposes, some drivers would question the decision based upon the test, especially if the outcome is negative.

There are a number of limitations to this project. The findings are based upon small samples, and the experimental sample used in the validity study may not be fully representative of the population for which the SDSA is being considered – namely the post-stroke population who are applying to return to driving. Most of the participants were self-referrals who volunteered after reading about the study in the Stroke Association's newsletter. The time post stroke tended to be longer for these participants than those recruited from hospital so the time since stroke is longer than in other studies and could explain some of the differences in findings. The time difference also means some participants' recall about their driving experience prior to their stroke may have been inaccurate. However, the fact people volunteered to take part because they were thinking about returning to driving, three or more years after their stroke, perhaps indicates that there is a demand for advice at a later stage of recovery than anticipated when the test was developed. Another implication of a large proportion of participants self-referring has been a large amount of missing data about the participants' strokes because doctors of self-referrers were less likely to complete the medical questionnaire. There was no incentive to do so if they were not part of the project. This meant we could not determine the severity and characteristics of all the participants' strokes and take this into account in the analyses.

It took a considerable amount of time to recruit the participants and level of recruitment was disappointing. The validity study was conducted over two years. During this time changes to staff needed to be made and the lengthy data collection period led to the use of multiple driving assessors. The majority of on-road driving assessments were conducted by one driving assessor (44), with the others assessing between one and 13 drivers. Consequently, the administration of the tests, scoring and judgements may not have been fully consistent over the entire duration of the project, although training was provided to promote consistency between assessors.

The validity and reliability of the on-road driving assessment as a predictor of fitness to drive was not assessed by this study and has not been established. This is partly because of a lack of measures to validate the assessment against. The on-road assessment has, however, been used by other studies as a criterion for fitness to drive and for this study it was used to assess the SDSA as a gatekeeper to an on-road driving assessment. Further research, which followed clients who undergo the assessment and collected data on their accident involvement could be undertaken.

It should be noted that the findings of the survey of SDSA users and inter-rater reliability study may have been biased by the sampling approach. Occupational therapists were sent questionnaires directly through NANOT, whilst other health professionals were contacted via the British Society of Rehabilitation Medicine and an advert on the British Psychological Society website. The advert appears to have been less successful as few clinical psychologists were identified as SDSA users or responded to the survey: the majority of survey respondents were occupational therapists. The use of the SDSA by clinical psychologists may have been different from that of the survey respondents and inter-reliability raters (who were recruited from the survey respondents) as they are more familiar with using psychometric tests.

Overall, this study has highlighted a number of concerns about the validity and reliability of the SDSA. Despite being relatively easy to administer and score, the tool is not always administered in a standardised way or interpreted consistently. The validity study has indicated that the predictive validity is inadequate for the tool to be used as a gatekeeper to the on-road driving assessment since in many cases would deny clients access to a driving assessment that they would otherwise pass. An adjustment to the pass-fail cut off might reduce the number of incorrect failures to a more acceptable level while still retaining a useful ability to identify and screen out drivers who would fail the road test. However, a larger dataset would be needed to establish this.

The current level of specificity and PPV may be considered sufficient if the implications of failure were less severe. If a driver who failed the SDSA was referred to a driving assessment centre after 3 months rather than being reassessed again using the SDSA, a PPV of 70% or less may be acceptable. Drivers who fail the SDSA would have their access to an on-road driving assessment delayed rather than denied. If a driver has to pass the SDSA before being referred, a driver who is incorrectly failed could be repeatedly denied access to an on-road driving assessment for some months; in this situation the current sensitivity and PPV are probably too low.

These findings are not to say that the SDSA should be abandoned completely. It may well be that use of the SDSA in combination with other cognitive and physical test will provide additional information to help practitioners decide who to refer for a driving assessment at a mobility centre. The survey of SDSA users highlighted that many users already use the tool in this way. The behaviour of the client whilst undertaking the test could also be noted as this may help the practitioner assess the person’s abilities.

The revised SDSA manual (Lincoln et al., 2004) includes a number of improvements but may benefit from another review to ensure practitioners have sufficient information to administer and interpret the tests in a standardised way. Training could also be provided to users to ensure they understand how to administer the tests.

Revising the predictive equations was beyond the remit of this study and the sample size was judged to be insufficient to both generate and then validate new equations. To overcome this difficulty, the data from this study could be combined with data from previous studies. The feasibility of doing this could be explored. A more valid measure than the UFOV also needs to found to establish concurrent validity of the SDSA.

The validity of a combination of SDSA, UFOV and other tests as a way of predicting the outcome of an on-road driving assessment would also be worthwhile exploring.

6 Acknowledgements

The authors would like to thank colleagues at TRL for their contributions to the project including Chris Baughan, who was TRL's Technical Referee for the project and provided comments on an earlier draft of this report, Lorna Pearce, Catherine Inwood, Kate Whitely, Su Buttress and Maureen Keigan. The assistance of Professor Nadina Lincoln, Nottingham University, who commented on an earlier draft of the report and took part in the inter-rater reliability study, was also invaluable. The authors would also like to thank the driving assessors Nick Goodrich, Charles Borsley, Edward Core, John Manning and John Sylvester and Yvonne Brown at MAVIS for their valuable contributions to the project,. The authors would also like to thank the clinics who referred clients to the study, and members of the project's steering group:

Dr Delyth Sheppard, DVLA Medical Branch

Elizabeth White, College of Occupational Therapists

Barbara Hatton, Wrightington Mobility Centre

Dr Gavin Newby, the Stroke Association

Mike McCaugherty, Different Strokes

for their co-operation and support for the project. Finally, the authors would like to thank all the participants of the various studies for taking part.

7 References

Ball, K.K., Beard, B.L., Roenker, D.L., Miller, R.L. and Griggs, D.S. (1988) "Age and visual search: Expanding the useful field of view". *Journal of the Optical Society of America*, 5, pp 2210-2219.

Ball, K., Owsley, C., Sloane, M.D., Roenker, D.L., Bruni, J.R. (1993) "Visual attention problems as predictors of vehicle accidents in older drivers". *Investigative Ophthalmology and Visual Science*, 34, pp 3110-3123.

Brown, Y. (2000) *Driving assessment following stroke*. Unpublished MSc Thesis. London: Brunel University.

Christie, N., Savill, T., Grayson, G., Ellison, B., Newby, G. and Tyerman, A. (2001) *The assessment of fitness to drive after brain injury or illness.* Crowthorne: TRL Limited.

Grayson, G., Stone, J. and Elliott, M.A. (unpublished) *Dementia and driving: An exploratory study.* Unpublished project report PR SE/292/01. Crowthorne: TRL Limited. (Available on direct personal application only).

Lincoln, N.B, Radford, K.A. and Nouri, F.M. (2004) *Stroke Drivers Screening Assessment. Revised Instruction Manual 2004.* Available on request from N.B. Lincoln, University of Nottingham.

Lundberg, C., Caneman, G., Samuelsson, S-V., Hakamies-Blomqvist, L. and Almkvist, O., (2003) "The assessment of fitness to drive after a stroke: The Nordic Stroke Driver Screening Assessment", *Scandinavian Journal or Psychology*, 44, pp 23-30.

Mant, J., Wade, D., Winner, S. (undated) *Stroke*. Available at http://hcna.radcliffe-oxford.com/stroke.htm. Last accessed March 2005.

Nouri, F.M. and Fanthome, Y. (1994) "Reliability of the Stroke Drivers Screening Assessment". *Clinical Rehabilitation*, 9, pp 157-160.

Nouri, F.M. and Lincoln, N.B. (1993) "Predicting driving performance after a stroke". *British Medical Journal*, 307, pp 482-483.

Nouri, F.M. and Lincoln, N.B. (undated) *Stroke Driver Screening Assessment. Instruction booklet*. Supplied as part of the SDSA Toolkit. Nottingham: Nottingham Rehab Ltd.

Radford, K.A and Lincoln, N.B., Hawkins, K.S., Murray-Leslie, C., Melly, S., Lilley, S.R. and Gorman, W.P. (2004) Validation of the Stroke Drivers Screening Assessment (SDSA). *Clinical Rehabilitation,* 13: 531-532.

Sackley, C.M. and Lincoln, N.B. (1996) Physiotherapy treatment for stroke patients: a survey of current practice. *Physiotherapy Theory and Practice*, 12: 87-96.

Sentinella, J., Stothart, G. and Keating, S. (unpublished a) *A survey of users of the Stroke Drivers Screening Assessment.* Unpublished project report PR SE/694/03. Crowthorne: TRL Limited. (Available on direct personal application only).

Sentinella, J., Sexton, B and Inwood, C. (unpublished b) *Inter-rater reliability of the Stroke Drivers Screening Assessment*. Unpublished project report PR SE/899/03. Crowthorne: TRL Limited. (Available on direct personal application only).

Walker, C., Walker, M and Sunderland A. (2003) Dressing after a stroke: A survey of current occupational therapy practice. *British Journal of Occupational Therapy*, 66 (6): 263-268.

Appendix A: List of variable names

Variable	Test	Measure
1. DDecision	Driving assessment	Decision making
2. DDivAtt	Driving assessment	Divided attention
3. DSpatial	Driving assessment	Spatial ability
4. DInteract	Driving assessment	Interaction with other road users
5. DInfoPro	Driving assessment	Information processing
6. DSelAtt	Driving assessment	Selective attention
7. DRVScore	Driving assessment	Mean total driving assessment score
8. SDot-Time	SDSA	Dot cancellation test: Time
9. SDot-Errors	SDSA	Dot cancellation test: Errors
10. SDot-False	SDSA	Dot cancellation test: False positives
11. SSqDirection	SDSA	Square matrix directions test
12. SSqCompass	SDSA	Square matrix compass test
13. SRdSignRec	SDSA	Road sign recognition test
14. SDSA Outcome score	SDSA	Difference between pass and fail equation scores (pass equation score – fail equation score)
15. UFOV risk category	UFOV	UFOV risk category
16. Usubtest1	UFOV	Sub-test1: Processing speed
17. Usubtest2	UFOV	Sub-test2: Divided attention
18. Usubtest3	UFOV	Sub-test3: Selective attention