University of Southampton

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
School of Civil Engineering and the Environment

# The Effects of Ageing on Driving Related Performance

by

**Muhammad Tariq Khan**

Thesis for the degree of Doctor of Philosophy

August 2009

UNIVERSITY OF SOUTHAMPTON

# ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
SCHOOL OF CIVIL ENGINEERING AND THE ENVIRONMENT

Doctor of Philosophy

## THE EFFECTS OF AGEING ON DRIVING RELATED PERFORMANCE

By Muhammad Tariq Khan

According to one estimate, about 40 percent of the driving population will be over the age of 60 by the year 2020 in the UK and currently, several hundred thousand drivers with dementia hold driving licenses. The number of motor vehicle crashes per unit distance of automobile travel is "U"-shaped, with risk increasing slightly between the ages of 55 and 60, but risk increasing with each successive five-year interval. Some individuals who have mild dementia possess sufficient driving skills to be designated as fit drivers. The most challenging assessment and decision for the physician/licensing authority as regards fitness to drive lies in drivers who are questionably demented or are in a state of very mild dementia.

In the absence of a reliable standard protocol, some clinicians make judgment based on self-reporting, which has risks associated with it as lack of insight and judgment are potential common traits of the population experiencing cognitive decline. Seldom is recourse made by health professionals to on-road assessment as a first alternative as it requires a fee and such testing centers are not readily available everywhere. This research addresses this issue of the identification of cognitive tests that can be used to assess an individual's ability to drive and especially of those individuals that are questionably demented and are the most difficult to identify. A younger and an older group consisting of 56 drivers in total were administered nine different cognitive tests and two drives (Drive-I and Drive-II) on the STISIM driving simulator. The cognitive test *ufov3* (involving the identification of a central target and simultaneously the radial localization of a peripheral target embedded in distracter triangles), which is the third subtest of the UFOV (Useful Field of View) test showed the highest discriminating ability in separating "poor-drivers" from "not-poor-drivers", with 92.86 % of the drivers correctly classified. The next best discriminating ability in decreasing order of strength was that of dichotic listening test, trail making test, rey-copy test and paper folding test. Also, age was found to be an excellent discriminator of "poor-drivers" and "not-poor-drivers" with 91.07 % of the drivers correctly classified. A composite cognitive measure consisting of the sum of all nine cognitive tests was not a better predictor than the *ufov3* test alone; overall it was still an excellent discriminator, classifying 89.29 % of drivers correctly. The commonly recommended Clock Drawing test and the Trail Making test did not emerge as significant predictors of driving ability. A general driving skills linear model for prediction purposes was derived that explained 59 % of the variation in a general driving performance index with the ufov3 test, the dichotic listening test and the rey-recall test as significant predictors. Recommendations are made as to how this test should be used to screen potentially at risk drivers.

Thesis Supervisor: Prof. Mike McDonald

## LIST OF CONTENTS

# List of Figures

# List of Tables

# DECLARATION OF AUTHORSHIP

I, **Muhammad Tariq Khan**, declare that the thesis entitled:
**The Effects of Ageing on Driving Related Performance**, and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself or jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

**Signed:** …………………………………………………………………………….

**Date:** ……………………………………………………………………………….

# Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Mike McDonald for all the support and guidance he provided throughout this study and for all the confidence he installed in me with his kind words. It has truly been a privilege to have him as a mentor. Heartfelt thanks are also extended to Dr. Pengjun Zheng for his continuous support and constructive comments. I would also like to thank Dr. Ben Waterson for the valuable knowledge he provided in statistics.

I would like to express my sincere appreciation to Mr. Brian Mould for assembling the STISIM driving simulator hardware and installing the software, to Mr. Doug Robinson and Mr. Eamonn Kennard for their invaluable comments during the refinement of the simulation drives and their enormous help in scheduling and facilitating the testing of the sample of drivers and to all the staff at TRG especially Melanie Hallford with her help in administrative matters. I would also like to thank all the subjects who participated in the study. Thanks are also extended to all my colleagues throughout my time at the University of Southampton.

I would like to express my most heartfelt thanks to my late father and to my mother (alive) for all the support and love I have received from them through out my educational life. Many thanks are also extended to my wife and my son who have endured the agony of separation through out the period of this research. I would like to express my utmost appreciation to my parents-in-law for their endless support and patronage.

# Abbreviations Used

| | |
|---|---|
| **AD** | Alzheimer’s Disease |
| **ADL** | Activities of Daily Living |
| **AIC** | Akaike Information Criterion |
| **BIC** | Bayes Information Criterion |
| **CD** | Compact Disc |
| **CDIS** | Clock Drawing Interpretation Scale |
| **CDR** | Clinical Dementia Rating |
| **DA** | Divided Attention |
| **DAT** | Dementia of the Alzheimer’s Type |
| **DPI** | Driving Performance Index |
| **DVLA** | Driver and Vehicle Licensing Agency |
| **DVLC** | Driver Vehicle Licensing Centre |
| **EM** | Expectation Maximization |
| **GP** | General Practitioner |
| **HC** | Hierarchical Clustering |
| **IADL** | Instrumental Activities of Daily Living |
| **IQ** | Intelligence Quotient |
| **LOO** | Leave One Out |
| **LTIT** | Landmark and Traffic Sign Identification Task |
| **MCI** | Mild Cognitive Impairment |
| **ML** | Maximum Likelihood |
| **MMSE** | Mini-Mental Status Examination |
| **MVA** | Motor Vehicle Administration |
| **NHTSA** | National Highway Traffic Safety Administration |
| **PASAT** | Paced Auditory Serial-Addition Task |
| **PC** | Personal Computer |
| **PDE** | Previously Defined Events |
| **RFT** | Route Following Task |
| **ROC** | Receiver Operating Characteristic |

| | |
|---|---|
| **SAS** | Simulator Adaptation Syndrome |
| **SDL** | Scenario Definition Language |
| **TBI** | Traumatic Brain Injury |
| **TLC** | Time-To-Line Crossing |
| **TTC** | Time to Collision |
| **UFOV** | Useful Field of View |
| **VD** | Vascular Dementia |
| **VIF** | Variance Inflation Factor |
| **WURT** | Washington University Road Test |

# 1 Introduction

## 1.1 Introduction

The proportion of licensed older drivers is increasing in the general population and a substantial proportion within this population group are experiencing a cognitive decline in functions that are critical to the driving task. Extended longevity has resulted in an increase in the number of people who at some stage acquire a degenerative brain disease. These individuals desire to continue driving as long as it is possible. Similarly, the survival rate of people recovering from an acquired brain injury has increased due to advances in medical technology and most of them wish to return to driving once they have surpassed the acute stage in their trauma. Considering the group performance of older drivers we see that there is a general gradual deterioration in driving performance with concomitant gradual increase in crash risk. However the individual performances of this group of drivers do not follow the same trend by virtue of inherent variability in the performance of this group. For example, there may be an older driver who shows signs of decline in his early 50s and by the time when he reaches the age of 60, he may be totally unable to safely function as a driver. On the other hand, another driver may continue to possess sound functional skills that may enable him to continue with the driving task without any apparent difficulty until a much later age. Although, older drivers show high accident statistics on a per-mile basis, most older drivers can drive safely. It is possible that the higher accident statistics of older drivers are on account of the cognitively impaired proportion of this population group because such cognitively impaired drivers (having some degree of cognitive impairment) have not been screened and are still driving. Older healthy subjects perform the driving task at a level that is comparable with healthy young adults. Therefore we may ask whether elderly drives contribute equally to the increased risk of accidents or is it that some proportion of this population group are at increased risk.

The difference between the direct effects of normal ageing and that of abnormal ageing (dementing disease especially in the early stage) relevant to driving skills, is less than clear-cut. It is possible that subjects exhibiting subtle cognitive changes may in fact have transgressed into the early stage of Dementia. Some individuals who have mild dementia possess sufficient driving skills to be designated as fit drivers; however, a stage /time will come when their cognitive impairment will increase and will ultimately render them unfit drivers. The most challenging assessment and decision for the physician / licensing authority as regards fitness to drive lies in drivers who are questionably demented or are in a state of very mild dementia. The prerequisite for driving cessation should be impaired competence for driving rather than a diagnosis of dementia since there is evidence that not all drivers in the early stages of dementia are incompetent drivers; the driving abilities of mildly demented/mildly cognitive impaired drivers have to be assessed, because no global ruling can be passed relevant to their condition.

There are certain medical conditions that have a tendency to bring about cognitive impairment to the extent that safe operation of motor vehicles is not possible and increasingly, such medical conditions have started to afflict people at relatively early ages. Even a small but significant number of younger people suffer from dementia who are likely to drive a motor car.

Vehicle testing is carried out in most European countries to ascertain performance of the vehicle as it ages but no recourse is made to testing elderly drivers; in the United Kingdom, more emphasis is placed on self-declaration of illness (to the DVLA) i.e., license holders are required under the law to inform the DVLA of any health condition that may influence their ability to drive. Subsequently, it is the responsibility of the DVLA Medical Branch to make judgement regarding a person's fitness to drive. After the DVLA is notified, individuals are required to fill and complete medical information forms. Medical advisors from the DVLA review and scrutinize these forms and a number of strategies are decided to carry out medical review. These may include: (a) requesting additional information from the GP/Consultant, (b) referral for specialist clinical assessment, or (c) acquiring an independent medical opinion. Some times as the DVLA deems necessary, it may require individuals to re-take the standard driving test or may refer him or her to a specialist driving assessment centre. Sometimes the family doctor/GP on his own initiative may contact the Medical Advisory Branch of the DVLA. However, this strategy only works if the family doctor is fully aware of all the disabilities/medical conditions and his client's driving habits. The situation is even more critical in cases where drivers suffer from dementia (dementia patients may lack insight into their illness), which is difficult to diagnose in its early stages and so many family doctors may be unaware of the condition. With regard to the issue of neuropsychological tests, there is little consensus on which tests can be used to predict driving safety and also which cognitive functions / domains are more relevant to driving that will assist more reliably in decision making regarding a persons fitness to drive. Hence there is no standard testing protocol (that is reliable) for assessing a person's fitness to drive after the onset of neurological or other illness. Therefore, different neuropsychological tests tapping different cognitive domains are in use. Hence, in the absence of a standard reliable protocol, the decisions regarding fitness to drive, that are based on the neuropsychological tests are doubtful and exude a low level of confidence on part of the clinicians/professionals. Due to the lack of a reliable standard protocol, some clinicians make their judgements based on self-report, which has risks associated with it as lack of insight and judgment are potential common traits of this population group. In this context, Christie et al (2001b) while carrying out a survey of clinical psychologists with regard to neuropsychological settings in the UK in assessing fitness to drive after head injury, observed: *"Overall, clinicians' decisions about a client's fitness to drive seem to be based on an eclectic approach with considerable reliance on "clinical impression.""* Also, the authors observed that a general comment by most respondents was that the criterion for opinion on fitness to drive have not been clarified by the DVLA except in cases of major mental afflictions/epilepsy. Seldom is recourse made by health professionals to driving assessment as a first alternative as it requires a fee and such centres are not readily available every where. Thus there exists a need for more information on assessment of fitness to drive with regard to neuropsychological tests, since medical information alone is not sufficient to assist in decision making of fitness to drive. This will also alleviate the need for the requirement of an on-road evaluation/assessment or can be a supplementary tool in addition to on-road assessment and will instill more confidence in decision making on part of the clinician.

Analysts disagree as to which tests are best predictors of driving and at which stage driving should cease. There is no single neuropsychological test that can reliably and

economically separate safe older drivers from those that are distinctly unsafe by identifying all deficits that are crucial to driving. Even the on-road driving tests may not identify important driving deficits in older drivers. Also, we see that there is in general, an absence of a theory driven choice of neuropsychological tests. There are plentiful neuropsychological tests that are not sensitive and specific enough to tap the neuropsychological constructs /cognitive domains relevant to driving. A comprehensive test battery should be able to tap multiple measures of each cognitive domain or construct so that the construct is more fully assessed relevant to the task at hand and also that structural or behavioural characteristic peculiar to an individual or group are avoided. The "test all" cognitive skills approach adopted with regard to neuropsychological tests does not bear relevance/resemblance to the actual driving task while assessing fitness to drive. We see that more emphasis has been placed on attributes that apparently seem important to the driving task like reaction time and general intelligence tests, while as little conceptual understanding has been developed of the cognitive skills that are required of driving. Higher-order cognitive skills are more important relevant to the driving task. Studies regarding neuropsychological testing of drivers sometimes have, and sometimes have not, showed significant correlations with performance results from a driving simulator/on-road test. These studies although had certain findings/conclusion relevant to a group but these could not be extended at the individual level and so could not be used to predict the outcome of a driving test for a particular individual. The lack of consistency and sufficient correlation between neuropsychological tests and driving skill can be attributed to the methodological disparity and maladoption of a logical test.
Although, the probability that the performance of older drivers deteriorates increases as their age increases, there is considerable variability in their performance profiles. Hence, we can not single out a specific chronological age at which a driver should be denied a driver's license.

## 1.2 Objectives

1. To identify a series of effective neuropsychological tests which tap cognitive domains/constructs that have significance for the driving task.

2. To determine the relationship between different driving performance parameters measured on the simulator and neuropsychological tests.
3. To develop a methodology to categorize driver conditions as passed/failed or their categorization (based on driving skill).
4. To model the relationship between categorization of drivers and neuropsychological tests.
5. To recommend a process whereby potential at risk drivers could be screened for further detailed evaluation.

## 1.3 Methodology (Overview)

In evaluating experienced older drivers, it was necessary to identify decline in their driving competence. This could be best achieved by comparing their driving performance with that of normal experienced drivers (control group), which would enable the identification of driving errors that were only present in the group that experiences the decline. Therefore a comparative approach was used.

Participants were recruited through advertisement leaflets that were distributed in the Southampton area including a number of bowling clubs. In carrying out this research, a group of older drivers and a control group consisting of young age drivers (normal experienced drivers) were given a battery of neuropsychological tests (cognitive tests) and then were made to drive two drives (Drive-I and Drive-II) on the STISIM driving simulator. The age requirement for the younger subjects was from 26-40 years, because statistically this is the safest age group and their performance is less likely to be confounded by the effects of age. Also this younger-group-age constraint ensured that participants would have at least 5 years of driving experience. The old age group drivers were above the age of 60 years. All drivers were to hold a valid UK driver's license with at least 5 years of driving experience. Subjects between the ages of 41 to 59 were excluded to minimize cross-sectional overlap in functionality. All drivers were to have adequate visual, hearing, communication and physical capabilities to complete the simulator driving tests/assessment.

A schematic diagram of the methodology adopted in carrying out this research is shown in Figure 1.1. The participants were first given a short (3 to 4 minutes) run on the practice drive (the beginning portion had S-curves, which can expose drivers prone to simulation sickness) to ensure that the driver was not prone to simulation sickness syndrome.

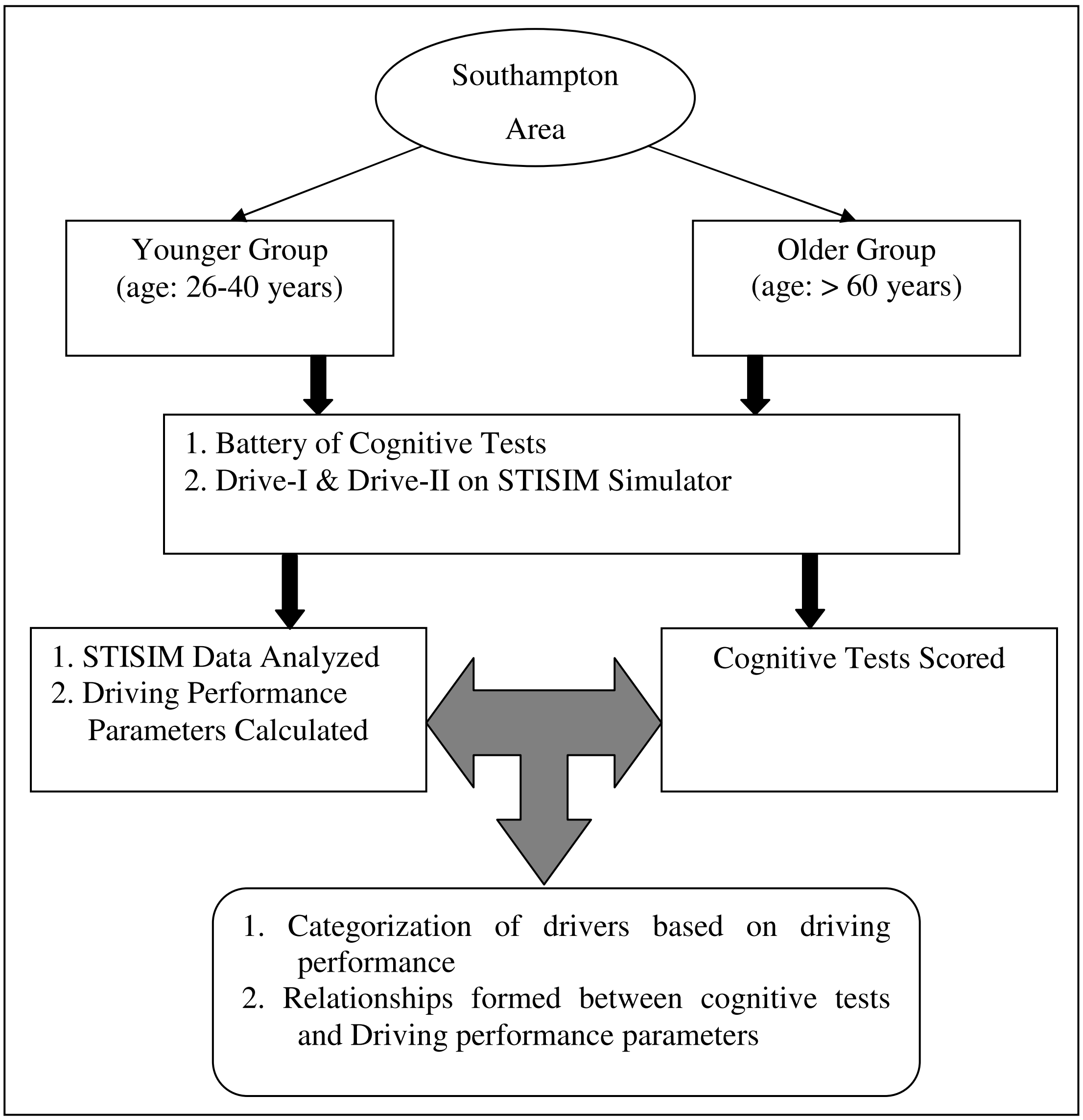



**Figure 1.1** Schematic diagram showing the methodology of research

Nausea, disorientation and ocular problems such as eyestrain, blurred vision and eye fatigue have been reported as some of the indicators of simulation sickness in fixed-base simulators (Mourant & Thattacherry, 2000). If a participant experienced the syndrome, the practice drive was immediately terminated and the participant was deemed unfit to take the simulation drive. If a participant did not feel any discomfort in the practice drive, then the rest of the protocol followed.

Firstly, participants filled out a brief questionnaire (except part IV which related to post-simulation issues), then they were given the following neuropsychological tests in random order: (1) Ufov Test (2) Dichotic Test (3) Trail-Making Test (4) Rey-Osterrieth Test (5) Paper Folding Test (6) Clock Drawing Test. This was followed by a practice drive for the Main Drive (Drive-I) and then the Main Drive. This procedure was repeated for the DA and Car-Following Drive (Drive-II). Finally part IV of the questionnaire was filled out. Numerous driving performance parameters were calculated from the STISIM generated data for further detailed analysis. Statistical techniques were used to extract useful information from the data and were subsequently used to categorize drivers based on driving performance using normal-mixture-model cluster analysis. The neuropsychological tests were scored and then multiple linear regression and logistic regression techniques used to form relationships between the neuropsychological tests and the measures of driving performance.

# 2 Literature Review

## 2.1 Introduction

The objectives for this chapter were to carry out a detailed literature review relevant to the effects of ageing on driving related performance and covered a broad and diverse range of issues that were relevant to the topic. In depth reviews were carried out incorporating the headings: (1) Accident Characteristics (2) Perception, Cognitive & Other Factors (3) Driving & Alzheimer's Disease/ other Diseases / other factors (4) Neuropsychological testing & Driving (5) Old Age & Mild Dementia and Driving (6) Driving Simulator & On-road tests.

According to a United Nations report, the median age of the world's population is increasing mainly because of a decline in fertility and a 20-year increase in the average life span during the second half of the 20th century (United Nations, 2002 cited in JAMA, 2003). These factors, combined with elevated fertility in many countries during the two decades after World War II (i.e., giving rise to the 'Baby Boom' generation), will result in an increase in the numbers of persons aged greater than (or equal to) 65 years during 2010 to 2030. Also, world wide, the average life span is expected to increase another 10 years by 2050. The largest increases in absolute numbers of older persons will occur in developing countries. During 2000 to 2030, the number of persons in developing countries aged greater than (or equal to) 65 years is projected to almost triple, from approximately 249 million in 2000 to an estimated 690 million in 2030 (U.S. Census Bureau cited in JAMA, 2003), and the developing countries' share of the world's population aged greater than (or equal to) 65 years is projected to increase from 59 percent to 71 percent (Kinsella, 2001).

It is reckoned that in the United Kingdom, twenty percent of the population is over 60, and by the year 2031, it would have increased to 30 percent (OPCS, 1992). As the proportion of elderly people is increasing in society, they are undergoing a change in their lifestyle and expectations. In future, older drivers will be driving greater distances and making more trips than today's older drivers (OECD, 2001). Importantly, more are currently drivers and expect to continue to drive into their old age. Also, the difference between the numbers of men and women drivers amongst the older generations is gradually reducing; according to a Department of Transport review ( Department of Transport, 2001), amongst 80-89 year olds in 1993, 44% of men and 11% of women held a current driving license. The DVLC (Driver Vehicle Licensing Centre) estimates that these figures are to rise to 65% and 35%respectively over the next 15 years. Currently, drivers over the age of 70 in the UK are over two million, which are expected to rise to four and a half million by 2015 (Noble, 2000 cited in Department of Transport, 2001).

Over the last 30 years, the biggest increase in active license holders has been among older women. For example, there was a 200 % increase in male drivers over 65 and a 600 % increase in female drivers over 65 between 1965 and 1985 as compared with 29% and 290 % increases in drivers (male and female) aged 17-59. In 1985, 57 % of men over 65 years of age and 14% of women over 65 held a valid driving license. By 1990, 55% of men and 20% of women over 75 years held valid driving licenses (Oxley, 1991 cited in Department of Transport, 2001). Thus, the difference in numbers of older men and women drivers is diminishing rapidly.

Similarly, in the United States, older adults represent the fastest growing segment of the population. This change, which sometimes demographers refer to as the "squaring of the pyramid," is altering the population structure from one in which many young people are at the base and few old people are at the top of the pyramid to one that resembles a rectangle and has an even distribution at all age groups (TRB Special Report 218, 1988 a). According to U.S Census estimates, the number of persons aged 65 and older will grow by 20 million, or 60 percent , over the next quarter century. Numbers of senior age drivers will grow at an even faster rate, as more elderly females and minorities are licensed and as those already having licenses maintain them longer. In the United States one in every seven licensed drivers is aged 65 or older; and by 2030, this number will approach one in five (Stutts et. al., 1998).

As the population ages, it is expected that the elderly will drive their personal automobiles more frequently and for longer distances, and will continue to do so until an older age than the elderly have done in the past. This includes not only the older driver (65 and older), but also the very old driver (those drivers 80 years of age and older). As the proportion of older drivers in the population increases, the burden of motor vehicle collisions in older people is also likely to expand. This demographic shift will have significant implications for safety of older drivers and the public, as it is well known that the number of motor vehicle crashes per unit of distance driven is "U"-shaped (see Figure 2.1), with crash risk increasing slightly between the ages of 55 and 60, but greater increases in risk with each successive five-year interval (Insurance Institute for Highway Safety, 1992).

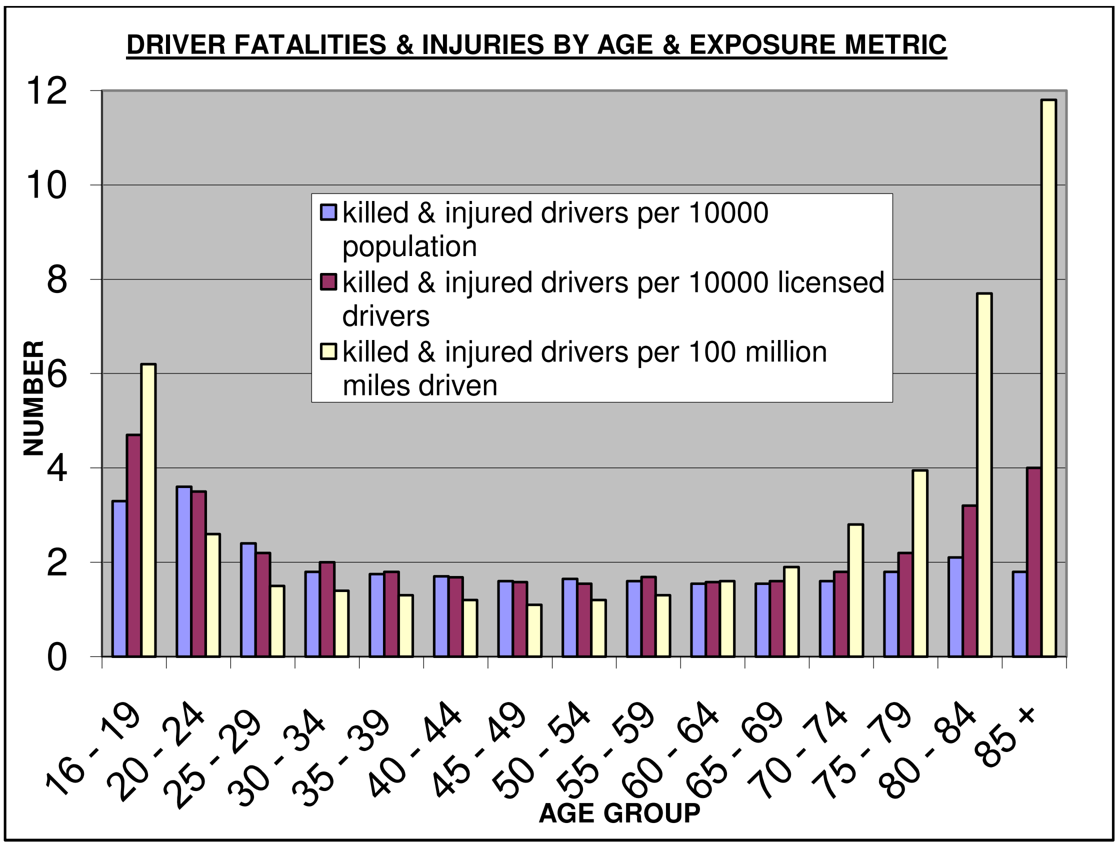



**Figure 2.1** Driver fatalities and injuries by age, related to population, number of driver licenses, and mileage driven (Hakamies-Blomqvist, 2004).

Similarly, the risk of involvement in crashes resulting in a fatality is also "U"-shaped relative to driver age (Insurance Institute for Highway Safety, 1992). According to a Transportation Research Board report (TRB Special Report 218, 1988 b), the proportion of traffic fatalities of the elderly is approximately equal to their proportion in the population.

An elderly person's risk of being killed or suffering a serious injury as a result of any road accident is between two and five times greater than that of a younger person because of their increased physical frailty; because as people age, their bone densities begin to decrease and also their immune system becomes weak and therefore, it becomes much more difficult for an older person to withstand the impact and injury from a motor-vehicle collision. When statistics based on all severities are examined, there is no age-related increase in total number of accidents for those over the age of 60 (DOT, 1997). No statistics claim that older people have anything like the number of accidents that young drivers have. For example, in 1996, 4253 drivers aged under 24 were killed or seriously injured on British roads, but only 1927 people aged over 60 were killed or seriously injured, and the rate per 100,000 population for all severities of accidents for the 60-69 age group was 117, whereas for the 20-29 age group, it was 507. However, statistics using population as an exposure metric show little increase with age, but when the same statistics are worked out by using miles driven as the exposure unit, the situation drastically changes as is evident from the "U"-shaped concept demonstrated earlier.

Accident statistics show a steep rise in per mile automobile accident risk beginning around age 65, with the fatality rate per million miles of travel being 17 times that of the 25-65 age group for those over the age of 65 (NHTSA, 1997 cited in McKnight & McKnight, 1999). As a further refinement of the break-up, the National Highway Traffic Safety Administration (NHTSA, 1994 cited in Rizzo et al., 2001), reckons that drivers aged 65 to 69 years are twice as likely to be involved in fatal multi-vehicle crashes as drivers aged 40 to 49 years, and drivers aged 85 years and older are 11 times as likely as drivers aged 40 to 49 years to be involved in such crashes. Drivers over the age of 65 years have the second highest fatality rate per mile travelled (the highest rate is held by drivers aged 15-24) even though they travel fewer miles per year than younger drivers. The fatality rate of drivers aged 80 years and older actually surpasses that of drivers younger than 24 years old. This increase is obvious as a function of greater likelihood of accident involvement and greater vulnerability to injury and death per involvement (Mackay, 1988). Evans (1991) has estimated that the fatality risk of older drivers grows for ages greater than 20 at an approximately uniform rate of about 2 percent per year. Department of transport (UK Department of Transport, 2001 cited in Hakamies et al., 2005) has stipulated that the increasing participation of elderly drivers in the future will lead to an increased number of serious accidents.

## 2.2 Accident Characteristics

The crashes involving older drivers more often occur in complex situations, where the driving task is not self-paced and there is a particular risk of cognitive overload. Compared with younger drivers, they are more likely to be involved in multi-vehicle collisions (as opposed to single-vehicle) in intersections and primarily are as a result of not obeying traffic rulers or traffic control devices and failure to yield right of way (Lundberg et al. , 1998; Caird & Hancock, 2002). In a typical intersection crash, the older driver is making a left-Turn (Right-turn in Britain) maneuver, when he is hit by an oncoming vehicle on the main road that has the right of way. As reported by Staplin & Lyles ( 1991), older drivers were more likely to turn left (Right in Britain) and collide with other drivers, but there was a less chance to be going straight and collide with left-turning (right-turning in Britain) vehicles. The high share of angle collisions, wherein the older drivers are hit from the side by an oncoming vehicle explains in part the serious outcomes of their accidents (Hakamies-Blomqvist, 2004). Past the age of 75, the risk of intersection collisions increases substantially for older drivers in almost all intersection maneuvers (Staplin & Lyles, 1991; Preusser et al., 1998). About 50 percent of all driver fatalities over the age of 80 are at junctions / intersection compared to only 23 percent for drivers under the age of 50 (Insurance Institute for Highway Safety, 2000 cited in Caird, 2005). Also, older Drivers also have been found to be the legally responsible ones in their crashes (Dulisse, 1997; Cooper, 1990). In Britain, seventeen to nineteen years olds had the highest accident rates, although older drivers (65 and older) had approximately twice the number of accidents involving failure to obey intersection control and far higher numbers involving turning across traffic ( Moore, et al., 1982 cited in Staplin & Lyles, 1991).

It has been observed that multi-vehicle and side impact crashes account for the majority of crashes sustained by older drivers and as compared to younger drivers, their propensity of involvement in side impact crashes is two fold (Adkins et al., 1999). Older drivers are also overrepresented while making maneuvers such as merging into traffic, changing lanes and backing up (Evans, 1991; Stamatiadis et al., 1991; Preusser et al., 1998; Staplin & Lyles, 1991). It was observed in a study by Mourant (Mourant, 1979 cited in Viano et al., 1990) that the accidents of older drivers involving alcohol, skidding or loss of control were appreciably lower than the average accident rates. McGwin and Brown (1999) while analyzing police crash data in the state of Alabama found out that alcohol involvement was much less of a factor in older driver crashes as compared to the crashes of younger drivers which involved a single vehicle, one or more driving errors and higher speeds. They also noted that the most common violation leading to crashes among crash-responsible older drivers was failure to yield. Other factors such as unseen objects/person/vehicle, failure to heed signs/signals, improper lane change and improper turns were also among the contributory circumstances as compared to those of middle-aged and young crash-responsible drivers. Crashes in adverse weather (e.g. rain, sleet, snow) were less common among older than either middle-aged or younger drivers. Also, older drivers were less likely to be involved in crashes caused due to driver fatigue, travelling at high speed, during the evening and early morning, on curved roads and involving a single vehicle. Failure to heed signs and grant right of way at intersections have also been cited by Kline et al. (1992) as critical actions manifested by older drivers.

In a study of the crash database from Kentucky from 1995 to 1999, Chandraratna & Stamatiadis (2003) using logistic regression also found out that relevant to limited-access highways, older drivers had higher crash involvement in high-speed lane changes. According to a Federal Highway Administration report, older drivers have an excess of turning and entering crashes at stop sign-controlled intersections (Knoblauch et al., 1995 cited in Preusser et al., 1998). For older drivers, careless or inaccurate lane changes, careless backing and driving the wrong way on one-way streets has also been reported (Mcknight, 1988; Yaksich, 1985 cited in Guerrier et al., 1999).

As reported by McKnight & Urquijo (1993), in a study conducted by the National Public Services Research Institute, specific behaviours were highlighted for the identification of deficient drivers, based on a 1000 sample of referral forms used by police in five states. The primary behaviours that brought drivers to the attention of police officers were:

- Driving the wrong way down a one-way street or on the wrong side of a two-way street.
- Failing to yield or stop.
- Leaving the roadway
- Turning across oncoming traffic.
- Slow speed
- Rear-ender
- Backing up
- Crossing lane markings
- Failing to yield to pedestrians or cyclists
- Miscellaneous/missed

These different accidents, violations and observations are summarized in Table 2.1. For example, driving the wrong way down a one-way street or on the wrong side of a two-way street contributed to many violations but few accidents.

**Table 2.1** Frequency of Behaviours contributing to Accidents, Violations, and Observations of Officers (McKnight & Urquijo, 1993).

| Behaviour | Accident | Violation | Observation | Total |
|---|---|---|---|---|
| Wrong way | 29 | 149 | 13 | 191 |
| No yield / stop | 74 | 114 | 3 | 191 |
| Off road | 176 | 8 | 1 | 185 |
| Turning across traffic | 46 | 43 | 0 | 89 |
| Slow speed | 0 | 56 | 9 | 65 |
| Rear-ender | 49 | 0 | 1 | 50 |
| Backing | 32 | 1 | 1 | 34 |
| Crossing lane marking | 5 | 25 | 0 | 30 |
| No yield to pedestrians or cyclist | 16 | 5 | 3 | 24 |
| Miscellaneous/missed | 58 | 43 | 39 | 140 |

### 2.2.1 Accident Details

The complexity of dynamic traffic interactions and traffic control information at intersections is indeed a problem for older drivers to reckon with. In typical intersection collisions, older drivers apparently fail to see the opposing vehicle in time or do not see at all and therefore, are unable to exercise an evasive maneuver (Hakamies-Blomqvist, 2004). Since intersections have high traffic and the visual situation is cluttered, therefore, there is a tendency on part of the older drivers to fail to detect critical roadway events because of the high information load. McKnight & McKnight (1999) reported that older drivers face considerable problems such as difficulty comprehending instructions, judgement of gaps, selection of appropriate speed and visual search, especially in traffic intensive situations.

According to De Raedt & Ponjaert-Kristofferson (2001), visual scanning, attention in the environs of the visual field, communication with other participants of the roadway system and violation of road sign are linked to the inefficient operations of elderly drivers at intersections. In an experiment to measure driver comprehension of Left-Turn signal and sign configuration, Drakopoulos and Lyles ( 1997) found out that comprehension of traffic signals deteriorates with driver age in correct answer rate and serious error rate. Consequently, older drivers who misunderstand as to the action they have to take may give up their right-of-way or violate another driver's right-of-way on a greater number of occasions as compared to younger drivers.

Misjudgment of the distance or speed of an oncoming vehicle or misjudgment of gaps is the main factor contributing to older drivers' higher involvement in intersection accidents (Stamatiadis et al., 1991). To execute a successful turning act at an intersection, an appropriate gap in the traffic stream must be located, vehicle shifted into the gap and accelerated so as to achieve the ambient speed of the traffic flow. A gap may be accepted or rejected, as perceived safe or unsafe by the driver; hence gap acceptance behaviour may be described as probabilistic in nature. The minimum time value is called the "critical" gap. There are several factors that govern the selection of an acceptable gap: vehicular flow, vehicle acceleration characteristics, rolling vs. complete stops and waiting time (e.g. long queue). Staplin & Lyles (1991) reported that older drivers have problems judging acceptable gaps and time-to-collision (drivers make an estimate of the time it takes moving at a constant speed to reach specified points in their path) and are further aggravated by older driver's generally slower response times. Hills and Johnson (1980) reported that younger drivers allowed a constant time gap and accordingly increased the distance at higher versus lower speed, whereas, older drivers maintained a constant distance gap. They also found out that relative to younger drivers, older drivers underestimate approaching vehicle speeds, thus overestimating arrival time of the oncoming vehicle which can have tragic consequences. Scialfa et al. (1991) showed that older drivers had a propensity for overestimating oncoming vehicle speed at lower speeds and underestimation at higher speeds, relative to younger drivers.

Staplin (1995) while making simulator and field measurements of driver age differences in left-turn gap judgements, reported that older drivers relied on perceived distance alone while making gap acceptance judgements during left turns (right turn in Britain). Older drivers' gap judgements did not change significantly for vehicles travelling at 48 km/hr and 96 km/hr, thus resulting in disproportionate risk for older driver when there is an isolated speeder in the opposing traffic stream. However, the young drivers allowed a constant time gap, thus catering for greater distance at higher speed versus lower speed. However, as pointed out by Hakamies-Blomqvist (1993) there could be an involvement of faulty estimation of the time needed for the driver's own actions / maneuver, keeping in view the long experience that older driver has had as a "normal" driver. In other words, the older drivers fails to take into account sufficiently their age-related slowing down, since these changes are slow and gradual and are therefore overseen which result in failure to change time estimates of their own actions in such forced-paced tasks. While investigating (multidisciplinary investigation) about fatal accidents of older drivers aged 65 or more in Finland in 1984 to 1989, Hakamise-Blomqvist (1993) observed:

*"............ the accidents of older drivers were very sudden and would not have been prevented by giving the older drivers a few hundred milliseconds more time, i.e. the amount corresponding roughly to the age-bound differences in psychomotor reaction times (Welford 1980; Summala and Koivisto 1990). Simple or even complex psychomotor reaction times do not reflect the augmented need of time at intersections correctly, and thus are not a*

*satisfactory basis for road design decisions, because they take into account only the slowing down of reactive action, not the slowing down of self-initiated complex behaviour."*

Older drivers also have shown tendency to fail to anticipate the movement of other vehicles in the intersection, thus leading to errors and conflicts (Treat et al., 1979 cited in Caird & Hancock, 2002). The significant differential between the speeds of older and younger drivers has been implicated by McGwin & Brown (1999) as a cause of dangerous accidents. Latency in response speed, erratic behavior / maneuvers and hesitancy of older drivers can create ambiguous scenarios for other drivers at intersections, which can prove catastrophic (Fozard et al., 1994; Hakamies-Blomqvist, 1994). Using in-car observations, a study performed by the Federal Highway Administration (Byington et al., 2001 cited in Chandraratna & Stamatiadis, 2003) noted the poor positioning adopted by older drivers while making left-turns, often failing to signal prior to making turns. While studying the crash database from Kentucky from 1995 to 1999, Chandraratna & Stamatiadis (2003) noted that high-speed lane changing was a problematic maneuver for older drivers and they often failed to detect vehicles in blind spots. A closer analysis of such maneuvers revealed that majority of such crashes were of the side-swipe type (as compared to rear end crashes). Involvement in angle crashes was non existent while they were changing lanes; this shows that their lane changing behaviour was non-aggressive as compared to younger drivers. Further, while analyzing the database, they found out that since a one-way road intersection possesses fewer conflict points, elderly drivers had lower likelihood for being involved in left-turn (right turn in Britain) crashes on one-way roads versus two-way roads.

In order to identify driver errors implicated in left-turn (right turn in Britain) accidents, Chovan et al. (Chovan et al., 1994 cited in Caird & Hancock, 2002) made an in-depth analysis of left-turns and came up with the following factors (exactly reproduced below):

- Inadequate slowing before entering the intersection.
- Failure to sense or comprehend traffic signals or signs.
- Failure to anticipate the actions or intentions of older drivers and pedestrians.
- Failure to detect an oncoming or crossing vehicle.
- Failure to accurately judge the gap or velocity of an oncoming vehicle.

- The physical masking of oncoming or crossing vehicles by other vehicles already in the intersection.
- Obscured line of vision by vehicles in front of the driver.
- Failure to coordinate travel within the intersection in accordance with the timing of the lights.
- Weather-related problems and road conditions.
- Roadway geometry limiting sight distances.
- Driver impairment due to loss of capacity, illness, drugs, fatigue, or alcohol.
- Improper signaling.
- Driver distraction by internal or external factors.
- Poor maneuver execution (i.e., too fast or slow, wrong trajectory)
- Driver impatience or hurry.

### 2.2.2 Accident (Additional Observations)

The speed of traffic and the rapidity with which dynamic changes take place at intersections can not be controlled through actions of the older drivers who fail in these situations. Accidents of older drivers are primarily linked to problems with visual scanning, attention in the periphery of the visual field, interaction with other road users, insights regarding traffic situation and violation of traffic control devices (De Raedt & Ponjaert-Kristoffersen, 2001). Hakamise-Blomqvist (1993), while investigating fatal accidents of older drivers aged 65 or more in Finland in 1984 to 1989 found out that accidents happened in a sudden manner from the drivers perspective; 44 percent of the elderly drivers had not anticipated danger prior to the accident compared with 26 percent of the younger drivers, primarily due to attention and perception errors.

As pointed out by Caird & Hancock (2002), a number of driver error categories have been designated as playing a role in the accident causation of older drivers. These include failures of perception, attention, cognition and action. Failures of attention translates to not seeing the other vehicle at all or seeing it too late such that it is too late to do anything about it (Rumar, 1990). Also, it has been found that the time required to search for a target or visual search performance worsens with age (Scialfa et al., 1999 cited in Caird

& Hancock, 2002) and obviously, this could have implications for intersections as performance in such situations is always under time pressure.

Although, accurate perception of distance to critical intersection features (e.g. Static features like islands, pedestals and other raised features etc) is necessary for the sound use of these facilities, transportation analysts give more credence to motion perception. Objects in our environs such as pedestrians and vehicles, are consistently changing position in time and space. Motion perception is the ability to perceive these changes relevant to the task at hand. In motion perception, dynamic stimuli (other moving vehicles) are the focus of attraction (Staplin et al., 1998). An estimate of the time taken to reach specific points in driver’s path when moving at constant speed, is termed time to collision (TTC) estimate. Time to collision estimates are hypothesized to be based on either an optic flow process or on a cognitive process (Staplin, 1995). In optic flow process, the driver’s analysis of the relative expansion rate of an image (e.g. an oncoming vehicle) over time provides the necessary cue to estimate TTC; in this process, the driver relies on two dimensional information and the angular separation cue (the image gets larger with the passage of time) provide the necessary means to estimate TTC. The cognitive process utilizes speed and distance information to estimate TTC and is essentially makes use of three dimensional information. As reported by Staplin et al. (1998), several studies (Schiff & Detwiler, 1979; Cavallo, et al., 1986) favour the optic flow modality which supports the concept of two-dimensional angular separation cues (relevant to background information) to estimate TTC. The optic flow theory has also been favoured by Lee (1976). As reported by Staplin (1995), a decline (possibly exponential) relevant to younger subjects has been noted in older subjects regarding the ability to detect angular movement. In fact, older drivers may require twice the rate of movement to get a idea that an object’s motion-in-depth is approaching, given a brief (2.0 seconds) duration of exposure to the scene (Staplin et al., 1998).

When making a left turn (right turn in Britain), motion in depth is in fact more challenging to the visual system, because expansion rate of the image of the vehicle on the head-on approach is utilized and not movement across the retina (in case the other vehicle was approaching at a right angle). The human perceptual system is more tailored to motion across the visual field versus motion in depth, which is always difficult to reckon with (Liebowitz, 1986 cited in Caird & Hancock, 2002). Schiff and Oldak (1990), have observed that estimation of arrival time is more accurate when vehicles approach at right angle to the driver. Also, detection of vehicles could be made more difficult by the low contrast between vehicles and the environmental background, however, both contrast sensitivity and acuity play an important role in road sign recognition (Owsley, 2004).

## 2.3 Perception, Cognitive & Other Factors

In order to operate a motor vehicle safely, multiple objects and events need to be continuously monitored so as to be pre-attentive as to where critical hazards may lie. Allocation of attention between onboard instruments and the roadway environment and coding of information from the perceptual (central and peripheral vision) and other senses has to be performed. The knowledge of road rules, awareness of spatial routes, vehicle positions and operations and sound decision-making and execution are most crucial. Prompt countermeasures need to be taken on feedback of safety errors and at the same time, achievement of travel goals, vehicle operation/integrity, signs of fatigue or incapacity need to be reviewed/monitored and appropriate strategies adopted ( Rizzo, et al., 2001). These function have been known to decline with ageing and neurologic disease, and can pose significant risk to safe driving operations.

Coordination and functioning of several cognitive processes is necessary for safe and efficient driving, these include perception, attention, memory (declarative, procedural and working) and executive functions (decision making and implementation). When these processes are impaired, as a result of neurologic or psychiatric diseases, the risk of driver error and motor vehicle crashes increases ( Rizzo & Dingus, 1996). Psychomotor abilities and general mobility are also important for the safe operation and control of a motor vehicle (Marottoli et al.,1994). Impaired decision making is a crucial factor in driver error that leads to vehicle crashes (Van Zomeren et al., 1987). Impairments in memory and in decision making can coexist independent of each other in the same individual. However, decision making will be affected if memory is impaired because memory enables the recall of all stored situational contingencies to deal with a specific traffic scenario, which will serve as input to the decision making process (Skaar et al. 2003 cited in Rizzo, 2004). Cognitively impaired drivers are less likely to realize their errors and thus are liable to make driving errors. Cognitive abilities / decrements are a determinant of driving behaviour and safety errors, which consequently predict accidents (Rizzo, 2004). In a study by Shinar (Shinar, 1978 cited by Parasuraman & Nestor, 1991), it was estimated that 25 to 50 percent of motor vehicle crashes result due to driver inattention. According to cognitive psychologists, there are at least three types of attention: selective, divided, and sustained attention (Parasuraman & Nestor, 1993) (see glossary). These three varieties of attention are not independent and separate entities and it has been shown that they are under the control of different but enmeshed *"networks of cortical and subcortical"* structures in the brain.

Rizzo et al. (2004), while elaborating on attention state:

*"Executive functions control our focus of attention (3). Focused attention is thought to permit consolidation of information temporarily stored in visual working memory. Without focused attention, we can be unaware of marked changes in an object or a scene as in "change blindness" (4-7), and traces of retinal images in visual working memory fade without being consciously perceived or remembered ("inattentional amnesia"). The very act of perceiving one item in a rapid series of images briefly inhibits ability to perceive another image, the "attentional blink" (8-10). These perceptual errors depend on interactions between attention and working memory and may increase with visual and cognitive decline associated with ageing, fatigue, medications, and neurological disease."*

The ability to engage in independent, purposeful, self-directed and self-serving behaviour can be described by Executive functioning. Executive functioning enables a person to focus and avoid distraction, do multitasking, have mental adaptability, plan and appreciate the future consequences of actions, self-monitor for errors and adjust behaviour accordingly and have abstract reasoning abilities (Hopewell, 2002). Being unaware of ones impaired cognitive defects is termed as Anosognosia. Anosognosia stems from deficits in executive function and it can adversely affect the functional effect of impairments in other cognitive domains (Anderson & Tranel, 1989 cited in Rizzo et al., 2001). Humans are single channel processors and therefore can not simultaneously properly attend to more than one item at a time; however, they have the capability to switch attention from one task to the other in the driving environment. As pointed out by Owsley et al.(Owsley et al., 1991 cited in Rizzo, el al., 2004), older drivers may experience a decrement in executive control and thus the ability to switch attention between tasks that are crucial to driving such as road terrain tracking, keeping track of the spatial change of vehicle locations, reading maps, signs, traffic lights and in-vehicle instruments, checking mirrors and switching attention between different sensory modalities. Kray and Lindenberger (Kray and Lindenberger, 2000 cited in Bieliauskas, 2005) stress that age negatively affects the ability to maintain and coordinate alternating tasks in working memory. McDowd & Filion (McDowd & Filion, 1992 cited in Bieliauskas, 2005) suggest that inhibiting attention to irrelevant stimuli is one of the primary difficulty that increases with age.

Automatic tasks in driving for experienced drivers consist of gear changing, lane keeping and steering etc., these are routine control tasks and are regarded as automatic compared with monitoring/scanning and decision making in dynamic traffic situations, which require higher order processing and the use of Working memory. Working memory is "….the ability to process information while maintaining intermediate products, goals, and associated strategies of processing online" (Naftali, 2000). The importance of working memory in driving cannot be overemphasized because, for example, while making a left turn (Right turn in Britain), based on the dynamic changing situation of oncoming traffic and driver's own vehicle capabilities (e.g. acceleration), information has to be processed /operated on, stored, retrieved and decision made (Guerrier et al., 1999). De Raedt & Ponjaert-Kristoffersen (2001) also confirmed that visuo-spatial function with working memory play an important role in left-turn (right turn in Britain) performance, primarily because the judgement of speed and distance of oncoming vehicles is involved. The situation is further exacerbated because multiple oncoming and crossing vehicles require that drivers coordinate their turn movements into gaps that are multiply constrained because of the directional split of traffic and that some times more than one lane is in each direction. Scialfa et al. (1994) in a study of age differences in the useful field of view concluded that decrements in search performance as a result of age are attributable to age related changes in eye movements, working memory and useful field of view (UFOV) (Scialfa et al., 1994 cited in Guerrier et al., 1999).

With regard to driving, a simple information processing model for understanding driver error is shown in Figure 2.2 below:

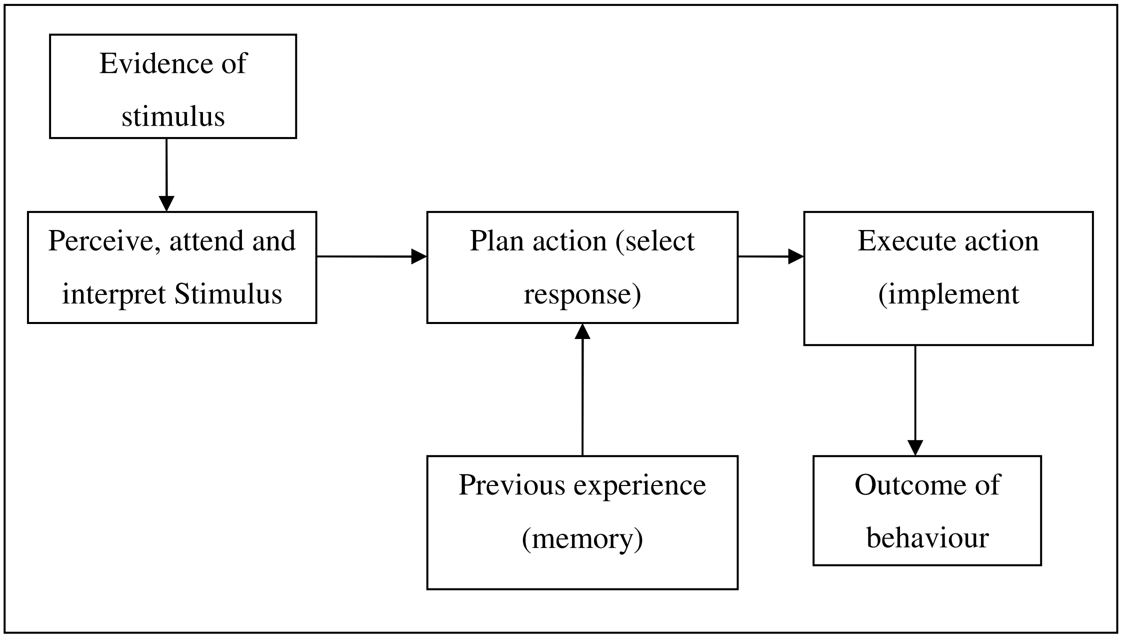



**Figure 2.2** Information Processing Model for Understanding driver Behaviour. (Rizzo, 2004).

The crashes of older drivers occur in complex situations where the traffic environment is not self paced and the risk of cognitive overload is high; therefore, it seems logical that age-associated cognitive overload is a primary logical factor in such crash causation (Hakamies-Blomqvist, 1996). Specifically, cognitive functions that support adequate visual processing and the coordination and integration of perception and motor skills are highly essential to the safe and efficient operation of motor vehicles in their environments. Many of the skills necessary for operation of a motor vehicle safely may be compromised with age or as a result of various medical conditions that often accompany ageing.

The cognitive functions deemed crucial to the driving task include attention, memory, scanning/visual spatial skills, information processing, rapid decision making and problem solving (Colsher and Wallace, 1993; Shinar, 1993 cited in Stutts et al., 1998). As pointed out by Owsley et al. (1998a), the fulfilment of the requirement of minimum visual acuity as per license requirements does not guarantee that an older driver will be safe on the road, because there are many other pertinent factors besides visual acuity that have significant ramifications for the safe control of motor vehicles (Owsley, et al., 1998a); such as visual field loss, contrast sensitivity deficits, visual attention impairment, cognitive impairment, cardiovascular disease, diabetes and medication usage. The use of certain medications such as sedatives, hypnotics, antihypertensives, antihistamines, anticonvulsants, antilipemics, hypoglycemic agents, pain medications and antidepressants has been known to be associated with accidents (Ray et al., 1993 cited in Rizzo, 2004). The detrimental effects of certain medications on driving is attributed in part to neurotransmitter systems involved in facilitating decision making and in working memory (Rizzo, 2004). Lundberg et al. (1997) and Ball et al.(1998) support the idea that in the crashes of older drivers, visual and cognitive decrements are the most important causal factors. Rizzo (2004) states that cerebral visual impairments in drivers makes them liable to *"look but not see"*, despite the fact that information load is low. This resembles the situation where neurologically normal operators perform under conditions of extreme fatigue (such as for example air traffic controllers during prolonged intensive monitoring of radar screens).

Since driving is a highly visual task, therefore, apparently, it has been assumed that the primary cause of the driving difficulty of the elderly is the presence of visual problems /eye diseases. Accordingly, most driving license issuing agencies have put a lot of emphasis on the assessment of visual acuity. However, as reported by Ball et al. (2006), although, several large scale sample studies have attempted but have failed to show a link between visual deficits (including several indices of visual function) and crash involvement, thus showing that visual function alone is a poor predictor of driving performance. Conventional measures of visual field assess visual-sensory sensitivity, whereas a test called the Useful field of view (UFOV) is linked to higher order processing skills, such as rapid visual-processing speed, selective and divided attention. In a study by Ball et al. (1993), to identify visual factors associated with increased vehicle crashes in elderly drivers, useful field of view test had high sensitivity (89 percent) and specificity (81 percent) in predicting the crash history of elderly drivers. It was observed that older drivers with significant deficits in useful field of view were six times more likely to be involved in accidents during the past 5 years. Significant correlation was obtained between crashes and eye health status, visual sensory function and chronological age, but these parameters were poor at distinguishing crash-involved drivers from crash-free drivers. In fact, According to the author, these tests (sensory tests, such as visual acuity and peripheral field sensitivity) do not reflect the visual complexity of the driving task, and are more relevant to clinical diagnoses and assessment of ocular disease/vision loss. The driving environment is quite complex, where vehicle control has to be negotiated in a cluttered environment through the simultaneous use of both central and peripheral vision to process both primary (high priority) and secondary (low priority) visual tasks and where the prediction of important events in time and space is unpredictable. Therefore, simple visual sensory tests fail to capture the visual demands of driving. In the Ball et al. (1993) study, the subjects whose visual acuity was better than 20/20, 43 percent had a useful field of view (UFOV) reduction of greater than 40 percent (the threshold amount) and 41 percent of the subjects who had a useful field of view (UFOV) reduction of greater than 40 percent (the threshold amount), showed an average loss of visual field sensitivity of less than 2.5 dB. Visual sensory and cognitive deficits can occur in older people together or separately. In another study (Ball et al., 1990 cited in Ball et al., 1993), it has also been shown that UFOV shrinkage can occur even in older subjects with excellent visual field sensitivity.

Owsley, et al.(1998a) carried out a prospective cohort study of 294 drivers with three years of follow up from 1990-1993 to identify whether measures of visual processing ability, including the Useful Field of View test, are related with crash involvement in older drivers. In their study, visual attention and visual processing speed was assessed using the Useful Field of View test. The Useful Field of View test is defined as (Owsley, et al., 1998a):

*"……….. the visual field area over which one can use rapidly presented visual information. Unlike conventional measures of visual field area, which assess visual sensory sensitivity, the useful filed of view test additionally relies on higher-order processing skills such as selective and divided attention and rapid processing speed. The test consists of a radial localization task in which a subject must identify the radial direction of a target (a silhouette of car) presented up to 30 degrees in the periphery, while simultaneously discriminating 2 targets presented in central vision ( a silhouette of a car versus a truck). By varying the eccentricity*

*of the peripheral target ( at 10 degrees, 20 degrees, or 30 degrees), the visual field area over which a subject can acquire information rapidly can be estimated. In some trials, the peripheral target is embedded in distracting stimuli. Thus, the task has both divided attention components (i.e., the subject must perform a central discriminating task at fixation while localizing a simultaneously presented target) and a selective attention component (i.e., the subject indicates the radial direction of the peripheral target even thought it is embedded in other discriminating stimuli in the periphery). Another variable manipulated is the duration of the test display, which varies from 40 to 240 milliseconds. Performance is expressed as a function of three variables: the minimum target duration required to perform the central discrimination task (subset 1), the ability to divide attention between central and peripheral tasks successfully (subset 2), and the ability to filter out distracting stimuli (subset 3). Performance in each of the 3 subtests is scaled from 0 to 30. In addition, performance in the 3 subtests is non independent because speed of processing is relevant to all 3 tests, and attention abilities are relevant to subtests 2 and 3. Performance in the overall useful field of view task is a composite score expressed as percent reduction (0 percent to 90 percent) of a maximum 30 degree field size (maximum field size of the test apparatus screen at the viewing distance). Using a previously established cut point, impaired useful field of view was defined as a 40 percent reduction or greater."*

The UFOV task depends on an individual's speed of processing, divided attention and selective attention performance and taps abilities that are vital to driving at the attentive (serial) and pre-attentive (parallel) levels. Stimulus and task features that are critical for driving are indeed incorporated in the useful field of view test. A driver with reduced UFOV may perform as if he or she has tunnel vision and yet he or she may not show any abnormality on standard vision perimetry tests, which place more emphasis on maximal estimates of sensory function vis-à-vis attention effects (Rizzo, 2004). Owsley, et al. (1998a) found out that older drivers with a reduction of useful field of view of 40 percent or greater were more than twice as likely to have experienced an accident. Reduction in useful field of view is quite prevalent in the older population. In fact, in a population based study, approximately one third of the older subjects had a 40 percent or greater reduction in useful field of view (Rubin et al., 1997 cited in Owsley, et al., 1998a). In another study, an overall correct classification rate of 85.4 percent (crashes versus no-crashes) was found  when UFOV was used as predictor variable in a logistic regression model (Goode et al., 1998 cited in De Raedt & Ponjaert-Kristoffersen, 2001). Sekuler et al. (2000) describes the deterioration in the useful field of view as a decrease in efficiency of extracting information from a cluttered scene. They also found out that the deterioration of UFOV starts early in life (by 20 years or younger).

In another study of older drivers performed in University of Alabama at Birmingham, Owsley et al. (1991) found that deficits in information processing ability as measured by the useful field of view test and deficits in cognitive abilities were related to crash involvement. By incorporating their parameters in a model they were able to explain 20 percent of variance in crash involvement. Further, it was reckoned that older drivers with poor scores on the UFOV or exhibiting poor cognitive status had 3-4 times more accidents (of any type) and 15 times more intersection crashes than subjects without those problems.

Chandraratna & Stamatiadis (2003) studying the problematic driving maneuvers of older drivers analyzed the crash database of Kentucky from 1995 to 1999 and found out that their high speed lane change crashes consisted 82 percent of the sideswipe type and 10 percent rear-end type. They further inferred that the side-swipe crashes reflected problems with peripheral vision and inattention and that the rear-end type depicted failure to correctly judge distance to the leading vehicle.

McKnight and McKnight (1999) contend that age dependant loss in abilities are interrelated such that association between any one ability and driving could be impacted by other abilities/variables. For example, the relationship between accidents and visual /perceptual deficiencies could be in part or completely due to cognitive deficiency in the same person. The effect of age on cognitive abilities is more pronounced than the effect of age on perceptual motor tasks. With age, peripheral factors influence reaction speed to a lesser extent than higher-order neurocognitive functions (Salthouse, 1985 cited in De Raedt & Ponjaert-Kristoffersen, 2001). For example Bieliauskas et al.(Bieliauskas et al., 1998 cited in Bieliauskas, 2005) compared the reaction time performance (to predict driving safety) of non-demented elderly drivers and drivers with dementia and found no significant difference. In old age, automatic routines remain relatively well preserved but older people find it very difficult to inhibit automatic processes in suddenly changing (and unexpected) situations (Rogers and Fisk, 1991). As pointed out by De Raedt &

Ponjaert-Kristoffersen (2001), such situations may be encountered in rear-end accidents when the leading vehicle suddenly stops;  the ability to switch from automatic to controlled processes plays a critical role in such situations.

The range of head and neck movements are important in order to check for vehicles, pedestrians and other obstacles at intersections and in the general roadway environment. When drivers approach intersections, distal and proximal visual information from many different directions is received through visual scanning using eye and head movement. The eyes actually make a series of little jumps while scanning the environment. The quick movements of the eyeballs from one spot to the next are called saccadic eye movements (pronounced "suh-cod-dik") ( Matlin, 2005). Saccadic eye movement bring the centre of the retina, known as the fovea, into position over object (e.g. vehicle or pedestrian), because the fovea has better visual acuity than other regions of the retina. Fixations, which allow the visual system to acquire information, occur during the period between saccadic movements. In order to accurately perceive information, the services of central vision have to be mobilized. Central vision occurs within a cone of vision in the vicinity of 3 degrees (where visual acuity is highest). Up to 10 degrees ( 10 degree cone), vision is fairly clear, beyond 10 degrees lies the region of peripheral vision, which may extend up to 160 degrees (160 degree cone) (Papacostas & Prevedouros, 2005). When peripheral cues enter the visual field, attentional processes determine which information is relevant for further detailed inspection using central vision. Central vision also facilitates the estimation of speed and distance judgement for gap selection. Isler et al. (1997), while making a study of age related effects of restricted head movements of drivers found that compared to subjects aged under thirty, the oldest subjects demonstrated an average decrease of about one-third of head movement in the right lateral plane, owing to physical neck conditions (e.g., arthritis, lower muscle tone, neurological disorders). When head movement is severely limited, virtually very little information is available from the head field, since upper torso movements are restricted by the car seat /belt. Hence the drivers would encounter blind angles in the traffic environment, from which relevant traffic information cannot be extracted. The use of larger eye movements (Saccades) to compensate for head movement restriction dose not work well as saccades have an amplitude of 15 degrees or less and if frequent larger eye movements are made, it puts considerable strain on the saccadic system (Bahill et al., 1975 cited in Isler et al., 1997). Also, drivers using corrected lenses may find that turning the eye beyond the limit of the lens may not prove useful. Carter et al. (Carter et al., 1983 cited in Isler et al.,1997) found out that longer saccadic latencies were exhibited by older people.

Many indictors of Physical health are based on the concept of functioning, like as to what extent, is the individual able to function normally and to carry out typical daily activities. Physical difficulties with instrumental activities of daily living may give indications that a person may be having problems with driving. Sims et al.(Sims et al, 1998 & Sims et al, 2000 cited in Owsley, 2004) reported that older drivers facing difficulty with performing such activities such as walking a mile, opening a jar, doing garden chores or light domestic work were involved in a crash or traffic violations.

Attention failures (especially at intersections) may result from visual search difficulties (McDowd & Shaw, 2000) improper division of attention (Ponds et al., 1988) and/or inappropriate selective attention (Ball & Owsley, 1991; Parasuraman & Nestor, 1991). An important attention failure on part of the driver may constitute as a result of inability to effectively detect/monitor changes in a busy intersection where the changes are rapid and the environment is dynamic. In this context, *Change blindness* plays a significant role. Change blindness refers to the inability to detect changes in an object or a scene (Simons & Levin, 1997 cited in Matlin, 2005), especially when people fail to notice a change in some part of a stimulus. According to O'Regan et al. (1999) change blindness results, when prominent changes are not noticed under natural viewing conditions because they occur at the same time when a brief visual interruption occurs such as an eye movement, a blink, a flicker or a camera cut in a film sequence. In fact, O'Regan et al. (1999) found that the phenomenon of change blindness can occur even when the change is not obscured or covered by the disruption. Dangerous events occurring in the driving scene can go unnoticed if they coincide with even very small apparently harmless disturbances. The variety of perceptual representations rapidly change from one glance to the next, as we drive along a busy street. If we precisely tracked each and every detail along the way, our visual system would be quickly overwhelmed by the trivial changes. Instead, our visual system is geared to forming accurate *"integrated gist"* or *"general* *interpretation"* of a scene (Matlin, 2005). As pointed out by Caird et al. (Simons & Levin, 1997; O'Regan et al., 1999 cited in Caird et al., 2005), recent research does not favour the long-held perspective that detailed and coherent picture-like representations of the world are stored by people from one view to the next. Instead, stable representation of a single object or its spatial location, is achieved through focused visual attention, which provides the necessary spatiotemporal coherence; therefore, as long as focused attention is provided to an object of interest or the region, visual representations may exist. Changes occurring in parts of a scene where attention is not focused will go unnoticed by viewers, primarily because there is no detailed representations of that part at that very moment. Intersections that have high flows and visual clutter will have a higher probability of missed changes (e.g. the appearance of a vehicle or a pedestrian from behind an initially occluding object) because drivers will fail to construct a coherent representation of the traffic scene (i.e. complete and accurate representation of each aspect of a visual scene), through focused attention. Caird et al. (2005), while using a modified flicker technique in order to induce change blindness so as to find the effects of time constraint on decision-making accuracy at intersections of young and older drivers, found that significantly more correct decisions (i.e. when change was detected) were made by young and middle-aged drivers compared to young-old and old-old. Finally, Caird et al. (2005) concluded that in particular, elderly drivers are more prone to missing important items at intersections thereby generating the typical "looked but did not see" (Cairney & Catchpole, 1996) errors.

Driving performance at intersections may also depend on the important construct of perceptual style. The extraction of salient information from a complex background is termed as perceptual style and field-independent are those people who most demonstrate this ability, while as field- dependent least demonstrate this ability (Witkin et al., 1962 cited in Mihal & Barret, 1976). Field-dependent people have more difficulty differentiating between relevant and irrelevant information compared to field-independent subjects. Field dependence has been shown to be related with slower recognition speeds for traffic signs (Lambert and Fleury, 1994 cited in Mihal and Barret, 1976) and Mihal and Barret (1976) found perceptual style to be related with crash involvement in a sample of 75 professional drivers from a utility company.

Perceptual, cognitive, physical, sensory and general driving knowledge deficiencies were the five main deficiencies identified in older drivers by Ballard et al.(Ballard et al., 1993 cited in Chandraratna & Stamatiadis, 2003). Specifically, cognitive functions like spatial orientation and perceptual speed have been found to decline in the normal course of ageing (Schaie, 1996). In the realm of sight, visual acuity, visual field, light sensitivity, night vision, color vision and spatial resolution are prone to decrements with advancing age (Ballard et al., 1993 cited in Chandraratna & Stamatiadis, 2003). Janke (Janke, 1994 cited in Preusser, et al., 1998) and Hu et al. (Hu et al., 1995 cited in Preusser, et al., 1998) are of the view that older drivers when confronted with divided attention tasks in visually cluttered environments where potential threats are coming from the periphery face considerable difficulty negotiating them. These hallmarks can usually be found at intersections.

## 2.4 Driving & Alzheimer's Disease / other Diseases/ other factors.

The matter of driver screening is no simple matter. The main problem lies in age-related diseases and impairments affecting driving skills rather than in effects of the ageing process per se. Since dementing illnesses are common in old age, a certain proportion of older drivers are in the early stages of a dementing illness or already clinically demented. Dementia is *"characterized by the development of multiple cognitive deficits…….. that are due to the direct physiological effects of a general medical condition, to the persisting effects of a substance, or to multiple etiologies"* (American Psychiatric Association, 1994). Alzheimer's disease (one of the common types of dementia), is a progressive age-related cognitive disorder associated with neurofibrillary tangles, extracellular plaques, and neuronal loss in the brain (Rizzo, 2004). Alzheimer disease may in fact be diagnosed in individuals as low an age as 50, but it is more prevalent in the old age group. O'Neill et al. (1992) report that there is a small but important number of younger people who suffer from dementia who drive and are more likely to carry passengers with them in a motor car. It is difficult to distinguish the cognitive deficits of normal ageing from that of mild stages of DAT (Parasuraman & Nestor, 1991) .The most common cause of abnormal cognitive decline in older adults is Alzheimer's disease (Cummings & Cole, 2002). Many individuals with early AD (Alzheimer's Disease) drive and their driving becomes impaired as the dementia progresses. Rizzo et al. (2005) reports that there is considerable evidence that progression of the disease begins years before it is clinically diagnosed.

In the very early stage of DAT (dementia of the Alzheimer's type), the cognitive decrements that prevail are not distinctly different from those that occur in normal older adults and hence the abilities that are pertinent to driving may be intact. Also, those mild DAT patients who in the beginning show language related problems, may have relatively sound driving abilities. Mild DAT patients who had had a robust cognitive constitution before the onslaught of the disease, may still have scores on standardized tests of memory and other cognitive functions that are within the bounds of normal functioning although compared to their premorbid (before the disease) condition, the decline would be still significant (Parasuraman & Nestor, 1991). In Alzheimer's disease, impairment in recent memory is one of the first cognitive impairment to surface. Hence driving tasks that depend on recent memory like identification of routes etc are likely to be more affected. However, memory aspects that do not depend on explicit recall, like procedural and implicit memory (such as basic operations of driving an automobile, gear shifting etc) tend to be maintained in mild DAT (dementia of the Alzheimer type) (Knopman & Nissen, 1987; Schacter, 1987 cited in Parasuraman & Nestor, 1991). The main issue of concern in driving in DAT is that the abilities that enables one to initiate driving are well preserved whereas as the abilities that make driving a goal directed and purposeful task are affected unfavourably (Parasuraman & Nestor, 1991). A number of compensatory strategies are employed by older drivers such as the adoption of lower speeds and avoiding difficult driving conditions (e.g. driving at night, in the rain or at busy traffic hours etc) (Hakamies-Blomqvist, 1994). However, in patients with dementia, some of these strategies may not work, for example by driving slowly or paying more attention to the traffic scene, an impairment in the efficiency in switching selective attention can not be compensated (Parasuraman & Nestor, 1991).

In one study, brain autopsies were performed on 98 older drivers who had died in motor vehicle crashes and it was found that 52 subjects (53 percent) had sufficient neuritic plagues to fulfill the standard neuropathological criteria to establish registry for Alzheimer's disease ( Johansson et al., 1997 cited in Rizzo et al., 2005). The matter is of quite concern due to the fact that none of these drivers were diagnosed as having AD and family members were often unaware of the problem and therefore, a fatal crash may sometimes be the first sign indicating AD. Since the cognitive decrement associated with normal ageing cannot be readily distinguished from that of very early stage dementia, and it is very difficult to diagnose the disease in the early stage, a considerable number of older drivers may continue driving because many of them will not be diagnosed as having the disease by physicians (Parasuraman & Nestor, 1993). Subjects with dementias are at a greater risk of becoming lost while driving and of being involved in automobile crashes.While examining dementia patients in their clinic, Lucas-Blaustein et al. (1988) found that 30 percent of them had been involved in accidents since the commencement of cognitive symptoms. The odds of them being involved in accidents are from 2.5 to 4.7 times that of age matched control group ( Reger et al., 2004). Retchin & Hillner (Retchin & Hillner, 1994 cited in Brown & Ott, 2004) report that the risk of crashes for elderly drivers with mild to moderate dementia is probably from 2- to 8-fold, compared to non-demented subjects. Alzheimer's disease (one of the dementias) affects 10 percent of individuals above the age of 65 and about 50 percent older than the age of 85 (Hebert et al., 2003 cited in Snyder, 2005). Carr (1997) reported that after the initial diagnosis, about 50 percent of the individuals with AD do not stop driving for at least three years. In a retirement community, Waller (1967) reported that 31 percent of the drivers were suffering from dementia. Compared to unimpaired drivers, drivers with dementia are less likely to report problems with driving and there is a mismatch between their actual driving performance and their perception of their driving skills and that of their caregivers (Staplin, et al., 1999). Also, they are less likely to limit their driving exposure to high risk /complex driving situation than drivers who have intact cognitive abilities, but have decrements in visual and physical abilities. Duration and severity of disease are important characteristics of dementia that have implications for driving ability. Drachman and collaborators (1993) after thorough examination of the issue found out that the risk of crashes rose above acceptable control rates beyond the third year of the disease. Alzheimer's disease is a neurodegenerative disease that is irreversible and is the most common cause of dementia in older subjects (Carr et al., 2006). The results of degenerative diseases (like AD) are progressive and ultimately, wider areas of the cerebral cortex and other regions of the brain are affected, whereas stroke and traumatic brain injury results in static brain lesions (Snyder, 2005). All dementias are progressive, but the rate of progression is not uniform from individual to individual or disease to disease. Alzheimer disease accounts for approximately half of all dementia cases. Features of dementia that help diagnose the disease include impairment of memory and at least one of the following cognitive derangements: aphasia, apraxia, agnosia, or a disturbance in executive functioning which lead to a significant decline from a previous level of functioning and result in impairment in social or occupational functioning (American Psychiatric Association, 1994). The rate of progression of a dementia (such as Alzheimer's disease) is quite heterogeneous in individuals afflicted with the disease. Therefore, instead of putting more emphasis on duration of disease, more emphasis has been placed on disease severity. One such measure that is employed is CDR (Clinical Dementia Rating). Dubinsky et al. (2000) reviewed literature with regard to driving ability and Alzheimer's disease status as determined by CDR rating. He used a conversion model to convert previous studies (that were not coded into CDR) into CDR rating categories and found out that relative to age-matched controls, subjects with AD (Alzheimer's disease) had a higher crash risk. In particular, those with CDR 1 dementia status had greater crash risk than those with CDR rating of 0.5. Thus as apparent, increasing dementia severity leads to deficient driving. Similarly, in a study of Alzheimer patients and healthy ageing people, significant relationship was found between Clinical Dementia Rating(CDR) and rating on a road test. Subjects having a CDR=0 (no dementia) were categorized as "safe" drivers (78 percent of CDR=0 subjects), compared to 67 percent of CDR=0.5 subjects (very mild dementia) and 41 percent of CDR=1 subjects (mild dementia). Only 3 percent of CDR=0 individuals were judged "unsafe" and 19 percent of CDR= 0.5 and 41 percent of CDR=1 individuals were judged "unsafe". The remaining subjects in each CDR group were classified under the category "marginal" (Hunt et al., 1997b; Hunt et al., 1997c cited in Staplin et al., 1999). Rizzo et al. (2001) tested older drivers with mild to moderate cognitive impairment due to Alzheimer's disease in simulated car crashes at intersections. The results showed that 6 of the 18 drivers with AD (33 percent) experienced crashes at the intersections compared to none of the 12 non-demented drivers of similar age. Hence most of the subjects with AD did not crash and fair control was exhibited by them. The authors suggested that some drivers with mild AD remain fit drivers and may remain safe on the road and therefore, they should be allowed to drive until such time when their cognitive impairment progresses to a state which is predictive of unsafe driving.

The failure to code, store and retrieve information in memory is the prime defect which occurs in early AD. Processing speed and attention abilities decline in AD and ageing, which can have profound effect on the information extracted per glance from a scene, useful field of view (UFOV) (it gets shrunk), and visual search abilities (e.g finding a familiar face in a crowd, operation of motor vehicle) (Rizzo et al., 2000). There is considerable variation in the extent to which different cognitive domains are affected in the early stages of the disease; however, all show signs of gradual progressive memory impairment. Performance on IQ, attention, verbal fluency, judgement, visuoconstruction and confrontation naming has also shown signs of decrements. Cognitive deficits in Alzheimer's disease and other dementias have serious implications for driving, since drivers are not able to tackle multiple stimuli simultaneously, maintain sustained attention, quickly respond to hazardous situations, judge distances and correctly interpret traffic control devices (Reger et al., 2004). Many drivers having Alzheimer's disease do not recognize their deficits and even if they do, they tend to down play their significance (Wild & Cotrell, 2003). Dementia patients often lack insight into their own behaviour and therefore are not good predictor of their own driving impairment. For example, Brown et al. (2005) in a study regarding prediction of on-road driving performance in patients with early Alzheimer's disease, examined 75 older adults and concluded that the neurologists' assessments were significantly related to on-road driving scores; however, the caregivers assessments were more valid than the self-assessments made by the dementia patients. Impaired awareness of one's compromised cognitive condition can result in the individual not taking steps to compensate for impairments and continue their driving routine despite the risk. Van Zomeren et al.(1988) conjecture that self-criticism and insight may be more crucial than the degree of cognitive deficits for a patient's fitness to drive. As cited in Rizzo et al. (2000), James (1890) has explained attention as:

*“ taking possession by the mind, in clear and vivid form, one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others ………”*

In Alzheimer’s disease, orienting attention, focusing attention and sustaining attention are at higher risk of showing decline. Therefore, where ever these attention constructs are employed, those tasks are liable to suffer e.g. tasks involving visual search, object recognition and working memory are particularly at risk (Posner, 1980; Mirsky et al., 1991; Perry & Hodges, 1999 cited in Rizzo et al., 2000). Rizzo et al. (2000) in a study conducted tests of attention and cognitive abilities on 42 individuals with mild AD and 22 controls subjects without the disease. They found out that divided attention, selective attention, sustained attention and visual processing speed (as measured by UFOV test) were the domains where AD patients performed significantly worse than the control group. Differences in age, education or basic visual function did not explain the differential performance of the two groups. Parasuraman (Parasuraman, 1998 cited in Bieliauskas, 2005) reports that decreases in the speed of saccades (eye movement), problems with shifts in attention and decrements in spatial scale of attention have been exhibited by DAT (Dementia of the Alzheimer’s type) patients.

Fitten et al. (1995) carried out a study involving two mild dementia (mild Alzheimer’s disease and mild Vascular dementia) and three age and health control groups in order to characterize on-the-road, behind-the-wheel driving skills and related laboratory performances/ neuropsychological tests of subjects with mild Alzheimer’s disease and Vascular dementia. The clinic control subjects consisted of 15 age-matched patients with diabetes (but without a history of stroke or dementia). Community controls consisted of 26 healthy age-matched older subjects (>60 years) and 16 young subjects (age 20 to 35 years). The drive scores of the mild AD group and the Mild Vascular dementia group were significantly different from the drive scores of the control group. The drive score performance of mild vascular dementia group was relatively better than that of mild AD group, and showed greater subject-to-subject variability than did the AD subjects. They explained this by reporting that vascular damage to brain areas that facilitate visual perception and attention occurs less frequently and consistently than neurodegeneration of AD. It was reported by Fitten et al. (1995) that on the road tests, DAT group drivers drove more slowly and committed more errors (e.g. driving into a street with entry prohibited).

In dementia, language disturbances occur commonly, but it has been noted that memory is always impaired. Decrements in memory (along with visuospatial impairments) may disorient a person and contribute to getting lost and consequently may result in the driver committing errors/violations, because intact short term memory enables a driver to retain information. Although language dose not have a direct impact on driving, it can influence strategic and tactical decision making (Lundberg et al., 1997). Due to the cumulative effects of diverse impairments, sometimes, a decline in two deficits in combination (e.g. cognition and vision) may be more detrimental to safe driving compared to a single deficit. Also, certain medications have adverse effects on driving ability. Using the Iowa driving simulator, Rizzo et al. (2001) found that accidents of subjects having AD were in fact related to visuospatial impairments; also, they were able to replicate on the driving simulator that some, but not all, subjects with dementia exhibit impairment in driving skill ( Rizzo et al., 2001; Rizzo et al., 1997). Cox et al. (Cox et al., 1998 cited in Brown & Ott, 2004) using an interactive driving simulator to assess the driving behaviour of Alzheimer patients relative to age matched controls found that Alzheimer patients displayed a propensity for driving slowly relative to the speed limit, driving off the road, took more time in negotiating a left turn (right turn in Britain) and applied less brake pressure when attempting a stopping manoeuvre.

Geographic disorientation, driving the wrong way on one-way streets or roundabouts and driving too slowly are some of the typical examples of impaired driving ability. Lundberg et al. (1997) state that subjects having dementia perform less well compared to elderly controls. However, there is considerable overlap between the scores of driving ability of the two groups and quite some subjects having mild to moderate dementia are bound to pass a standard road test. They further add that under these circumstances then, a standard driving test will not determine whether a subject with mild dementia will be safe in negotiating a critical / hazardous situation that may arise in traffic. Also, owing to the degenerative nature of the disease, a standard driving test cannot determine whether driving will be safe in the future. Hunt has identified a list of traffic situations /

manoeuvres that demented drivers find difficult negotiating (Hunt, 1994 cited in Staplin et al., 1999), these are:

- Left Turn (right turn in Britain) at Intersections: drivers fail to yield right of way or inappropriately interpret the traffic control devices with regard to the execution of left turns.
- Failure to Remember Routes: Even familiar and well travelled routes are not well remembered and may lead to the driver getting lost.
- Confusion in Selecting Pedals: Under stressful conditions or in emergency situations, drivers may press the gas pedal instead of the brake pedal or vice versa.
- Complex Driving Situations: in complex situations where the cognitive load is high and rapid cognitive processing /problem solving has to performed, the driver may stop in the middle of traffic or be unsuccessful in negotiating the traffic. Although, to an observer, stopping in traffic in that particular situation is not warranted.
- Failure of Interpretation: The drivers fail to correctly interpret or perform timely interpretation of verbal commands or instructions/suggestions from a fellow passenger, for the execution of an appropriate response.

Hunt et al. (Hunt et al., 1993 cited in Brown & Ott , 2004) compared the on-road driving performance of subjects having questionable to mild severe AD (Alzheimer's disease) with that of age matched controls. All subjects who had questionable or very mild dementia and the control group (normal subjects) passed the road test, while as 40 percent of the mildly demented subjects were deemed incompetent in the road test. In the mildly demented group deemed unfit to drive, poor scores was achieved on signalling from curb, attending to task, overall judgement, awareness of driving impact on others and driving at inappropriate speed. Also, general cognitive measure such as the CDR (Clinical Dementia Rating) and more specific tests of attention, visuoperception, language, memory and timed performance were correlated with driving performance. Joint mobility, coordination, strength or primary visual abilities did not correlate with driving performance. In a similar study that involved a larger sample, Hunt et al. (Hunt et al., 1997a cited in Brown & Ott , 2004) confirmed their previous observations that the severe the dementia (as judged by CDR ratings), the more impaired was the driving. In this

study, the percentage of control subjects, very mild dementia subjects and mild dementia subjects who failed the road test were 3 percent, 19 percent and 41 percent respectively.
ADLs (Activities of daily living) also play an important role in the context of driving ability. In one study, activity of daily living scores were related to driving ability whereas neuropsychological tests did not help identify those with bad driving (O'Neill et al., 1992). This makes intuitive sense since it is known that in many forms of dementing illnesses, aspects of procedural memory remain for a longer time than episodic memory. Like motor aspects of driving, many ADL functions come under the rubric of over learned procedures employing procedural memory and therefore, impairments in ADLs may indicate a more severe condition. Since, a mix of both skills i.e. over learned skills and ability to respond to novel/unexpected situations is required for the driving task, performance of IADL (Instrumental activities of daily living) tasks may better reflect performance on the driving task than, for example ADLs (Lundberg et al., 1997). IADL are more demanding and complex than ADLs (McDowell, 2006). IADLs includes activities such as financial management, use of transportation and shopping —— activities that are more complex and hierarchically arranged that integrate *"lower level ingrained habits and higher level planning and supervisory functions"* (Lundberg et al., 1997). In this context, Lundberg et al. (1997), also reports that typical crashes of demented drivers when analysed show lack of higher order skills more as a causal factor than deficiencies in basic vehicle handling. Carr et al.(Carr et al., 1990 cited in Fitten et al., 1995) reported in a study that some degree of cognitive impairment was prevalent in more than 60 percent of elderly drivers and that 25 percent of those drivers needed some help in bathing and dressing. However, Shua-Haim & Gross (1996) in a study of the driving ability of 41 patients with Alzheimer's disease found no correlation between driving performance and functional status evaluated by ADL and IADL. According to the authors, both ADL and IADL are general assessment tools for the elderly and are not specific for assessing the functional status of subjects with Alzheimer's disease or dementia.

Dobbs et al. (1998) using a comparative approach to identify unsafe older drivers assessed 155 older drivers with clinically significant declines in mental abilities on on-road test and compared them with the performance of a normal elderly control group (68 subjects) and a normal younger control group (30 subjects). They found out that hazardous errors were the single most important indicator of membership in the clinically impaired elderly group. More hazardous errors were committed by the cognitively impaired older drivers than by the two control groups, which had similar performance. Further detailed analysis showed that 50 percent of the hazardous errors occurred during lane changing, merging and approaching intersection manoeuvres; 21 perfect occurred during left turns (right turn in Britain), 15 percent while failing to stop, 6 percent in making right turns (left turn in Britain) and 8 percent in stopping manoeuvres.

There are certain medical conditions that have a tendency to bring about cognitive impairment to the extent that they precludes the safe operation of motor vehicles and such medical conditions can occur at any age (Dobbs et al., 1998).There are other diseases besides Alzheimer's disease that impact driving ability e.g., there are other types of dementia like Vascular dementia, Frontotemporal dementia and Dementia with Lewy bodies. Crash risk is not indicated by the presence of a disease condition by itself, but by the extent of functional impairment caused by the disease in the individual. In this context, the list includes acute episodes such as seizure, syncope, or stroke, or chronic conditions such as cardiovascular disease (particularly when impaired cognition and fatigue are also present), neurological disease such as dementia, Parkinson's disease and multiple sclerosis or musculoskeletal diseases such as arthritis (Gilfillan & Schwartzberg, 2005).

Also, in the elderly population, the prevalence of multiple clinical co morbid conditions and psychosocial stressors/factors can contribute to functional impairment (and thus contribute to impairment in driving skills) independent from dementia or in addition to dementia. These multiple clinical disorders and the relevant treatments that the elderly under go can create complex pathophysiology that may have a negative impact on driving, since numerous medication can bring about mental status changes (Wang et al., 2003 cited in Snyder, 2005).

Stroke can afflict people at all ages; however, its occurrence is more common in old age. Stroke is a relatively common medical problem; as per one estimate, as of 2002, 4,600,000 stroke victims were living in the United States (American Heart association, 2002) and the number of drivers who have had a stroke is increasing because of the ageing trends and demographics in United States (JAMA, 2003 cited in Uc et al., 2005). In the United Kingdom, each year as many as 5200 people of driving age suffer a critical traumatic brain injury (TBI) (Barnes et al., 1998 cited in Radford et al., 2004) and during the rehabilitation process, the important question of whether the patient is fit enough to resume driving has to be addressed since there are many cognitive impairments, which remain. Many of the drivers inflicted with TBI (traumatic brain injury) fail to realize their responsibility of informing the licensing authority and resume driving without appropriate evaluation or relevant advice (Christie et al., 2001a; Fisk et al., 1998; Pidikiti & Novack, 1991 cited in Radford et al., 2004). Stroke causes focal brain lesions which can cause problems with navigating around places i.e. topographical disorientation. The topographical disorientation can occur as a part of a more broader cognitive decline or it can occur in isolation. This can result in the patient getting lost and can result in other critical errors while driving. Uc et al. (2004) conducted a study of 32 participants with stroke and 104 neurologically healthy control subjects. The participants were given a route following task (RFT), which placed demands on driver memory, attention and perception. The RFT was similar to a real-world situation where a drivers follows verbal directions to get to a destination of interest. It was observed that drivers afflicted with stroke committed more navigational and safety errors than neurologically healthy drivers. The authors stated that executive functions, mental rotation of imagined space/image, recognition of landmarks from an altered perspective / unusual angle and comparison with mental model compiled from the set of initial verbal instructions, all play an important role in route navigation. As reported by Lundqvist (2001), impaired attention, decrements in cognitive processing speed and executive dysfuntion have been known to result from brain injury. When brain injury affects the frontal regions, it may result in impairment in inhibitory control and working memory, because, the frontal cortex controls attention and executive functions. In a study regarding stroke patients admitted to hospital, it was found that 76 percent had a perceptual deficit (Edmans & Lincoln, 1987 cited in Nouri et al., 1987) and therefore it is important that this crucial factor may be taken into account before a return to driving. Nouri et al. (1987) reported that two studies examined the effects of dysphasia (impairment of speech and verbal comprehension, especially when associated with brain injury) on driving ability but it was found that accurate prediction of driving skills was not provided by language performance alone.

According to Christie et al. (2001a), cognitive impairments in the constructs of memory, concentration, problem solving, decision making and general intellect have been associated with injuries or illnesses affecting the brain; increased aggression, which reflects changes in personality, may also occur. The ability to plan, monitor and self regulate behaviour (which comes under the rubric of executive functioning) and skills such as insight may show decrements in people who have traumatic brain injury (TBI), because of damage to the frontal lobe of the brain. Christie et al. (2001b) point out that individuals who have executive function deficits usually lack complete insight into their impairments. As a result of impaired insight, TBI patients may feel fit to drive, when in fact they have poor driving skills. Christie et al. (2001a) while carrying out a study in assessment of fitness to drive after brain injury or illness found that correlation between the driving instructor's judgement and the clinicians judgement was 0.8, while as the GP (General Practitioner's) judgement based on medical information had low correlation with the driving instructor's judgement. Therefore, it is apparent that in case of brain injured patients medical information alone is not sufficient to enable a decision regarding fitness to drive. According to Wilkinson et al. (Wilkinson et al., 1989 cited in Christie et al., 2001b), residual deficits in memory, attention / concentration, decision making and behavioural control are some of the long term repercussions of severe head injury. For very severe cases of TBI (traumatic brain injury) patients who had undergone extensive rehabilitation in hospitals, Brouwer & Withaar (1997) after reviewing studies relevant to fitness to drive, found that a re-licensing rate of slightly over 50 percent was present; while as driving was resumed without any difficulties by the less severe cases. After head injury, in many individuals, driving skills may be relatively intact. In this context, McKenna (McKenna, 1998) states:

*"The cognitive blueprints for carrying out the motor sequences of movements in driving depend on automatic* subconscious *areas of the brain. This is why many people ………. who have suffered a head injury…………. are often still able to drive safely and efficiently, at least as far as controlling the car is concerned."*

Visual sensory cues are affected in stroke (affecting various regions in the brain) which may lead to attentional decline. Also, the ability to recognize landmarks and traffic signs is affected, which provide important information relevant to a driver's route and about upcoming hazardous situations and safety regulations (Uc et al., 2005). When stroke stricken cognitively impaired drivers visually search for roadside targets, it can puts strain on their limited cognitive resources and can bring about decrements/impairment in driving performance. In this context, Uc et al. made a study in stroke stricken drivers, to assess their ability for visual search, road-side target recognition and their safety errors during a landmark and traffic sign identification task (LTIT). It was found that stroke drivers who were previously familiar with the area of town where LTIT was carried out committed a similar number of at-fault safety errors as the controls. Whereas unfamiliar (with the area of town where LTIT was carried out) stroke victims performed more safety errors than unfamiliar controls. However, the performance of a subgroup of stroke stricken drivers was well on all LTIT and some made no safety errors. The authors finally concluded that some stroke stricken drivers have intact driving skills and may be allowed to continue to drive.

The prerequisite for driving cessation should be impaired competence for driving rather than a diagnosis of dementia. This fact is also reinforced by the statement of Alzheimer's Association Board of Directors: *"A Diagnosis of Alzheimer's disease is not, on its own, a sufficient reason to withdraw driving privileges. The determining factor in withdrawing driving privileges should be an individual's driving ability. When the individual poses a serious risk to self or others, driving privileges must be withdrawn."* (Alzheimer's Association, 2001). Brown & Ott (2004) report that organizations such as the American Psychiatric Association, the American Academy of Neurology, the International Consensus Conference on Dementia and Driving, the Canadian Consensus Conference on Dementia and the combined group consisting of the American Association of Geriatric Psychiatry, the Alzheimer's Association and the American Geriatrics Society all recommend driving cessation for individuals with moderate to advanced dementia because in these states, there is sufficient cognitive impairment to be detrimental to driving. However, unanimity of decision with regard to patients having mild dementia does not exist. Brown & Ott (2004) further report that the majority of guidelines acknowledge the fact that in mildly demented drivers, driving abilities must be assessed, because no global ruling can be applied. They further elaborate that certain organizations even recommend probing the driving skills of drivers who have a history of traffic crashes or executive / judgement dysfunction. For individuals that have moderate to severe dementia, there is strong consensus that they should not drive, however decisions regarding those having mild dementia are problematic (Johansson & Lundberg, 1997). Brown & Ott (2004) also report that there is evidence to support that not all persons in the early stages of dementia are incompetent drivers. A group of researchers was invited by the Swedish National Road Administration in order to formulate a consensus of the issue of driving and dementia. The researchers (Lundberg et al., 1997) reached consensus that in case of moderate to severe dementia, driving should be avoided and certain individuals with mild dementia should be considered for specialized assessment with regard to driving competence. They also stipulated that a periodic follow-up is sufficient for mildly impaired non-demented drivers who have a stable and acceptable functional level and who do not exhibit any evidence of driving impairment. Carr et al.(2006) suggest that repeat testing at six- to 12-month durations must be conducted because driving skills are likely to decline. They also report that some patients with dementia do not recognize their driving impairments (due to lack of insight) and therefore, resist efforts from family members to prevent them from driving. The American Academy of Neurology in its consensus statement on driving and dementia has recommended that patients with a severity rating of CDR=1 or greater should cease to drive, because they have a substantially increased accident rate and driving performance errors; patients having a severity rating of CDR=0.5, pose a significant traffic safety problem and in their case, the Academy recommends that the patient should be referred to a qualified examiner for assessment of driving performance. The recommendations further add that since dementia is a progressive disease, therefore, clinicians should reassess dementia severity and driving performance every 6 months (Dubinsky et al., 2006). However, one aspect is without doubt: owing to the progressive degenerative nature of the disease, there is a time, when ultimately all demented drivers become incompetent drivers.

## 2.5 Neuropsychological testing & Driving.

Rizzo et al. (2005) while studying impaired response implementation of older drivers with cognitive decline on a driving simulator, administered a variety of neuropsychological tests and concluded that neuropsychological tests predicted driver performance and error because negotiating driving scenarios/situations depends on multiple cognitive domains, just as neuropsychological tests do. The authors assessed older drivers by observing their response to an emergency vehicle (such as a police car) parked by the road side. In such situations, drivers are required to detect and recognize an object in peripheral vision, recognize the situation, select an appropriate response and execute a safety manoeuvre in order to reduce the potential for crash with the vehicle or running over people situated in the vicinity of the police car. They suggested that older drivers with cognitive decline show decrements in situation awareness or executive control with regard to response implementation at the level of choosing one of several possible learned evasive motor manoeuvres/actions.

To a large extent, driving is automatized and is mediated by mental schemata. Schemata is a pattern imposed on a complex reality or experience to assist in explaining it, mediate perception, or guide response. But when complex situations arise or an unexpected event occurs, then controlled processing is required. These shifts between automatic and controlled processing take place as per the demand of the traffic scenario and the experience of the driver. Controlled processing comes into play when routine reactions do not suffice and the complexity of the situation necessitates the use of attention controller, the central executive. Thus, efficient driving is achieved through directed driving behaviour and attention control. Traffic situations usually require action to be implemented within a short period of time. Time pressure can affect controlled processing, since information has to be held in working memory temporarily. Therefore, processing speed is dependent on working memory (Lundqvist, 2001).

The intrinsic primary function of neuropsychological tests is to assess psychological functions at a functional level, while on-road tests evaluate an individual for driving skills at an activity level. Sometimes, the two approaches form divergent assessments primarily because in driving activity, individuals’ adaptive/compensatory measures, experience and attitudes and motivation also play role in maintaining safety, which are not accounted for in the neuropsychological tests (Summala, 1997; Matthew et al., 2000 cited in Lundqvist, 2001). For example, despite neuropsychological impairment, drivers who are well aware of their dysfunction and have a long experience in driving, may adapt their driving behaviour.

Reger et al. (2004), while conducting meta-analysis of 27 primary studies to examine a relationship between neuropsychological functioning and driving ability for adults with dementia, found a significant relationship between neuropsychological functioning and driving ability as measured by on-road tests and non-road tests. They found out that decline in cognitive functioning is accompanied by concomitant decline in driving ability. However, when care givers report was used as a measure of driving ability, mixed results were obtained when the relationship between driving ability and cognitive functioning was analysed. Visuospatial skills and attention were deemed the most helpful in screening at risk drivers. Visuospatial skills enable a multitude of tasks including correct automobile positioning and manoeuvring on the road, judging distances and forecasting spatial position of vehicles or their evolution in the road environment. Selective, divided and sustained attention are essential for the detection of potential hazards, to negotiate competing stimuli at intersections and to maintain sustained vigilance on trips that are long (Lundberg et al, 1997). Reger et al. (2004) suggested that while assessing driving skills on the on-road test, variation in traffic and road conditions and subjective scoring increase the variability of on-road scores, hence weakening the strength of the relationship with neuropsychological test scores.

Assessment of mental status is carried out when fitness to drive is considered. As reported by Reger et al. (2004), Fox et al.(1996) found that 94.7 percent of Aged Care Assessment Teams used the Mini-Mental Status Examination (MMSE) for assessing cognitive functioning, when evaluating fitness to drive. However, Reger et al. (2004) from their meta-analysis of 27 primary studies found out that when control and demented subjects were included, mental status (as measured by MMSE) showed a moderate relationship to road test scores, but when control participants were excluded, the significant relation ship ceased to exist. They commented that this may be due to the fact that only middle to late stages of some dementias (including Alzheimer’s disease), may reliably show signs of mental status changes and therefore, these stages would reflect universal changes in driving ability. Therefore, by including subjects with Alzheimer's disease (who experience mental status changes) and also control, quite large correlations are observed. Brown & Ott (2004) also report that mixed results have been obtained regarding MMSE and driving impairment. Fitten et al. (Fitten et al.,1995 cited in Brown & Ott , 2004) report that correlation between driving scores and MMSE scores at the middle range of the scale may exist but less so in the higher range of MMSE (27/30 and higher). However, in two studies ( Fox et al., 1997; Trobe et al., 1996 cited in Brown & Ott , 2004), MMSE could not predict future traffic violations or future accidents, because the MMSE is a very brief general cognitive measure that places emphasis on orientation to time and place, language (naming, repetition, comprehension, reading, writing, copying), attention and calculation (serial 7s, spell "world" backward) and immediate and delayed recall (three words) ( Strauss, et al., 2006). Snellgrove (2005) reports that the MMSE does not cater to the assessment of executive functions. The Swedish National Road Administration invited a group of researchers to frame a consensus on the issue of driving and dementia; relevant to the MMSE, one proposal (Lundberg et al., 1997) called that cut-off scores on the MMSE must be considered as being relative, and should form only a small part of the basis of decision making about driving and should always be secondary to a clinical evaluation. In the report (Lundberg et al., 1997) an MMSE score ≤ 10 when accompanied by a diagnosis of dementia, warrants immediate cessation of driving, because it indicates a sufficiently low level of cognitive functioning. Other paradigms relevant to other score ranges of the MMSE were also forwarded, but not all experts agreed on the proposals.

Duchek et al. (1998) examined the relationship between visual attention measures and driving performance in healthy older adults and individuals with mild dementia of the Alzheimer type. Subjects were assessed on on-road driving trials and were administered computerized experimental visual attention tests. They found out that measures of selective attention (i.e. visual search performance) had better predictive power than psychometric performance tests. They inferred that many psychometric tests are not process-specific, but are rather reflective of general processes and hence these tests may tap general cognitive status, rather than skills necessary for driving. They finally concluded that selective attention is related to driving skills in the early stages of DAT (Dementia of the Alzheimer type).

In a study of 38 brain injured subjects, Korteling & Kaptein (1996), assessed participants on tests of perceptual speed, time estimation, tracking reaction, information processing and an on-road test. They found out that perceptual speed and time estimation were significantly related with driving performance. 29 brain damaged patients and controls were put to neuropsychological tests, simulator driving and on-road testing by Lundqvist et al. (1997). The neuropsychological tests correctly classified 80 percent of the patients. Based on the outcome of their findings, they recommended using neuropsychological tests that measure attention, information processing and executive functions. The predictive validity of clinical opinions and neuropsychological tests to the outcome of an on-road driving test was evaluated by Christie et al. (2001a) in 39 subjects with head injury or generalized brain damage. Logistic regression was used to identify 5 tests that were significant predictors of driving skill. These tests measured visual memory, executive abilities, spatial awareness and attention. On account of practice or due to the over-learned nature of the driving task, some brain damaged people have the ability to compensate for their deficits. A mismatch may result between neuropsychological tests (that measure psychological functions at a functional level) and driving performance, because besides cognitive functions, the role of experience, adaptability and motivation can not be downplayed in driving (Ranney, 1994; Lundqvist, 2001; Lundqvist & Ronnberg, 2001 cited in Radford et al., 2004). Van Zomeren et al. (1988) report that controlled attentional processes are impaired after head injury, while as automatic processes are not. Nouri et al. (1987), while investigating the relationship between cognitive ability and driving after stroke, in which 39 pre-stroke drivers were assessed using a battery of cognitive tests and an on-road test, finally concluded that cognitive tests which involve complex reasoning skills appear most highly related to driving performance.

Parasuraman et al. (1992) tested 15 patients with mild to moderate dementia of the Alzheimer type and 15 healthy, age matched controls in order to examine cue-directed shifts of attention for a letter-discrimination task. Valid, invalid and neutral cues were given (indicating probable target location) through the use of central arrows that preceded the display of every letter. It was found that the reaction times when valid cues were used did not differ between DAT (dementia of the Alzheimer type) subjects and the control group. However, the reaction times of DAT subjects were significantly greater than the control group, when invalid cues were used suggesting that in early DAT, subjects retain the ability to focus attention to a spatial location, but show impairments in the ability to switch or disengage attention. Parasuraman & Nestor (1993) report that the ability to switch attention is also correlated with driving performance in normal individuals.

Studies have been conducted in the past to explore the relationship between driving performance and experimental measures that gauge the efficiency of focusing and switching attention. For example, Kahneman et al. (1973) conducted a study in which a test of auditory selective attention (the Dichotic Listening Test) was given to 117 professional bus drivers aged 22 to 32. In the dichotic listening test given to the subjects, a series of digits and word were presented at the same time to each ear. The drivers were told to concentrate attention on one specific ear (i.e. one specific channel) and to report the digits appearing on that specific channel. In the second part, they were told that when they hear a "particular" tone, they have to switch attention to the other channel (other ear) and report digits from that channel (on 50 percent of the second parts, the same ear is the relevant channel). Essentially the test was used to gauge three categories of errors in the realm of selective attention:

1. Number of errors committed in reporting the digits from the relevant ear (channel) (omission errors). This measure indicated the extent to which there was inability to focus attention on the relevant channel.
2. The number of digits reported from the irrelevant channel (intrusion errors). This measure indicated a person's vulnerability to distraction.
3. Since 50 percent of the second parts involved switching to the other channel, thus the number of errors in (1) and (2) as explained above following the switch in relevant channel in the second part were coded as switching errors (switching errors).This measure indicated failure of the attention mechanism responsible for switching.

Correlation coefficients of 0.29 and 0.31 were obtained between the number of errors committed in reporting the digits from the relevant channel and the number of digits reported from the irrelevant channel respectively with vehicle crashes over a one year period. And a correlation coefficient of 0.37 was obtained between the number of switching errors and crashes over a one year period. Similar findings have been replicated in other studies. In an even earlier study by Gopher & Kahneman (1971), the dichotic listening test was applied to a highly pre-selected group of cadets (of high-performance aircraft) in the Israel Air Force and a significant correlation of 0.36 was found with a three-level criterion in pilot training. For predicting different criteria of proficiency in flying high-performance aircraft, the test had promising validity. Also, pilots of high-performance interceptor and attack aircraft had significantly better performance (on part two of the test) than pilots of transport and slower jet aircraft. They finally concluded that in driving / flying under normal conditions and in the second part of the dichotic listening test, the requirement to reorient attention is a common feature. The ability to switch attention as measured by the dichotic listening test is an indicator of overall performance since in the driving and flying tasks, the operators of the machines (car & plane) do not passively wait for orienting signals but have to rapidly switch attention between the stream of events taking place. Intuitively, one would think that how is it possible that the scores on the dichotic listening test (a test dependent on auditory modality) could be correlated with crashes, when driving is apparently a predominantly perceptual task. To elaborate on this aspect, Avolio et al. (1985) conducted a study in which the auditory selective attention test (the dichotic listening test) and a visual selective attention test (developed by the author on the model of the auditory selective attention test and constructed to equal as a visual counterpart of the auditory selective attention test), were given to seventy two drivers (aged from 28-59). The authors found out that all three categories of errors on the auditory selective attention test (dichotic listening test) were significantly correlated (in the predicted direction) with individual accident rates. However, in the visual selective attention test, only omission errors and switching errors were significantly correlated (in the predicted direction) with individual accident rate. The intercorrelations among the errors of the auditory selective attention test and the visual selective attention test were positive and significant. Also, it was observed that the correlations between the switching errors of the auditory selective attention test and the visual selective attention test had the highest correlations (among the test battery) with individual accident rate. Finally the authors concluded that since both measures of selective attention (i.e. both modality-specific measures i.e. auditory as well as visual) were correlated with each other and were also correlated with an external task (i.e. the individual accident rate), therefore both may be tapping a central cognitive construct / domain, that is modality-free. However, the most distinct differentiation between the successful and unsuccessful cadets was provided by the switching error category. Similarly, in this study also, the authors concluded that since there were high intercorrelations between all three categories of errors and the dichotic listening test predicted performance in the *"primarily visually loaded piloting task"*, therefore the dichotic listening test in essence was tapping a single central cognitive function that was modality free. Parasuraman & Nestor (1991) give a list of other studies (along with relevant data in tabular format) that show significant correlation starting from 0.3 to 0.4 between measures of driving performance (some driving index) and at least one measure of selective attention. One primary fact that surfaces from these studies is that the largest correlations were obtained from the switching error category of selective attention. In this context, Kahneman et al. (1973) at the conclusion of their study suggested that to reorient attention from an earlier state of attention to a channel/stimulus is more difficult than to initially apply focused attention from an uncommitted waiting state. Posner (Posner, 1980 cited in Parasuraman & Nestor, 1993) has described these attention activities as the *disengagement* and *engagement* of attention, respectively. Disengagement and reorienting of attention are the particular attention traits that have shown signs of impairment in patients with mild DAT (dementia of the Alzheimer's type), although, there is no marked decrement in the ability to focus attention in these individuals (Parasuraman et al., 1992).

For the detection of unpredictable or infrequent events for extended periods of time, it is necessary that sustained attention or vigilance be maintained. However, it has been noted that when driving has been maintained for a long period of time (e.g. sitting behind the steering wheel for many hours), decrements of attention occur and detection speed (of stimuli) and accuracy tend to be affected adversely (Davies & Parasuraman, 1982; Warm, 1984 cited in Parasuraman & Nestor, 1993). Since under conditions of extended driving, a person's vigilance level tends to decrease and so it would seem logical to propose that crashes may result under certain driving environments. However, a robust association has not been found despite the fact that numerous studies have been conducted (Parasuraman & Nestor, 1993) to prove such an association.

With regard to neuropsychological tests that facilitate in determining whether a dementia patient is safe to drive, Lundberg et al.(1997) stipulate that numerous neuropsychological tests are not sensitive and specific enough to tap the behavioural and cognitive features /domains that may affect driving. They further add that some studies have shown reasonably high correlations between driving competence and neuropsychological tests, however, results have seldom been replicated and may not be applicable to the clinical setting. Especially, problems in the generalizability of correlational findings occur because the correlations were derived from general population samples and are being applied to a target clinical population. Also, common neuropsychological tests fail to add extra predictive power over and above that is known on the basis of a subject's diagnostic status, because these tests are often part of diagnostic batteries. According to Lezak (Lezak, 1983 cited in Christie et al., 2001b), impairment in high level functions such as "executive" functioning is particularly difficult to detect, because situational inflexibility, perseveration, social disinhibition and impulsivity are masked by the motivating and structured context of formal neuropsychological testing. The driving task can be considered in terms of three levels of behaviour i.e. strategic, tactical and operational, but there is no provision in standard neuropsychological tests/probes to measure such attributes (Michon, 1985 cited in Rizzo et al., 1997 & Van Zomeren et al., 1987).

Gopher (1982) gave the Dichotic Listening Test to 2000 flight cadets in the Israeli Air Force and found switching errors had the highest correlation with success in flight training and the dichotic listening test did add to the predictive value of the test battery employed for pilot selection and concluded that since there were high intercorrelations between types of errors and the dichotic listening test predicted performance in the "primarily visually loaded piloting task", therefore the dichotic listening test in essence was tapping a single central cognitive function that was modality free. Kahneman et al. (1973) conducted a study in which a test of auditory selective attention (the dichotic listening test) and a brief form of Raven's Progressive Matrices Test (a short intelligence test) was given to 117 professional bus drivers aged 22 to 32. Finally, the authors reported that *"the validity of the selective attention test (dichotic listening test) was not* *due to differences in intelligence; the short intelligence test .........did not discriminate significantly between the criterion groups, and its correlation with the attention test (dichotic listening test) was low."* Also, Parasuraman & Nestor (1991) report that they conducted a study of older and younger drivers and found small and non-significant correlation between accident rate and Wechsler Adult Intelligence Scale (intelligent quotient). An ideal neuropsychological test should must have high validity and reliability, should possess good sensitivity and specificity, be simple and speedy to implement, and should not be expensive and be well tolerated by subjects (Fitten, 2003). In describing the general characteristics/attributes of neuropsychological tests, Ball et al. (2004) have highlighted that:(a) a single neuropsychological test does not reflect a pure measure of a single cognitive domain, (b) a single test only partially taps a specific domain, (c) more than one domain are tapped by almost all neuropsychological tests, (d) many of the measures obtained from different neuropsychological tests are highly related because the cognitive constructs themselves are interrelated.

Tallman (Tallman, 1992 cited in Reger et al., 2004) states: *"……Thus, what is (needed) are tests that are correlated with driving abilities within a mildly impaired group of individuals"*

## 2.6 Old Age & Mild Dementia and Driving

With regard to health, the elderly are considered a very heterogeneous group. Even in older people of the same age, there exists considerable variability in different attributes. After middle age, health deteriorates exponentially over a period of 1 to 3 decades. Some older people experience a quite rapid decline in health while others will have a slow decline and they will be afflicted with disabilities quite late in life (Fitten, 2003). Hence, global decline is rarely observed (Schaie, 1996). During the course of normal ageing, some cognitive functions e.g. spatial orientation and perceptual speed have been found to decline (Schaie, 1996). Older drivers are more sensitive to noise and hence in performing a task, they require stronger signals to react (i.e. they have a lower signal-to-noise ratio) (Lundberg, 2003). Also, older adults have problems related with the spatial and temporal integration of items into larger units, which may reflect their deficits in working memory

(Welford, 1985 cited in Lundberg, 2003). The onset of age-associated diseases affecting cognitive functions even further exacerbates the situation and thus increases the risk of motor vehicle crashes in elderly drivers. Compared with old age, in DAT, there is considerable loss of neuronal cell loss and also there are far greater number of neurofibrillary tangles (Morrison & Hof, 1997; Price et al., 1991 cited in Parasuraman et al., 2000).

Fitten et al. (1995) carried out a study involving two mild dementia and three age and health control groups in order to characterize on-the-road, behind-the-wheel driving skills and related laboratory performances/ neuropsychological tests of subjects with mild Alzheimer's disease and Vascular dementia. The clinic control subjects consisted of 15 age-matched patients with diabetes (but without a history of stroke or dementia). Community controls consisted of 26 healthy age-matched older subjects (>60 years) and 16 young subjects (age 20 to 35 years). There was no significant difference in the drive scores of the three control groups and the older healthy individuals performed at a level that was comparable to that of the young healthy adults, in a suburban type drive.

Although there is increased likelihood of cognitive impairment in older drivers, Withaar et al. (Withaar et al., 2000 cited in Bieliauskas, 2005) report that the results of neuropsychological tests show a wide variation (within this group) and hence translates into mild to moderate correlations with accidents. Bieliauskas (2005) reports that the effects on driving of normal ageing and abnormal ageing (dementing diseases) are not clear-cut. The results from past studies may have been wrongly interpreted, as in those studies, demented individuals may have been included in the apparently healthy old group, because they did not show clinical symptoms of dementia (but in fact had undergone delicate cognitive changes). This would have erroneously shown a general impairment in the driving skills of the apparently healthy group and would have falsely given the impression of a continuum between normal ageing and abnormal ageing (dementing diseases).

Daigneault et al. (Daigneault et al., 2002 cited in Bieliauskas, 2005) in a study of older drivers of age of 65 and above suggested that measures of executive function could help in the identification of those older drivers that are at significant risk of accidents and it is these drivers who fail to exercise compensatory behaviour in their driving protocol. Since the prefrontal cortex has important contribution in exercising executive control (Funahashi, 2001), therefore putting more emphasis on executive functioning is quite in harmony with the “frontal ageing hypothesis” which stipulates that in the ageing process, the prefrontal cortex is one of the most affected areas of the brain (Bieliauskas, 2005).

Driving is a day to day real-world example of a divided-attention task, because in driving, the driver has to coordinate different activities under both low traffic and high traffic situations. If the activity being performed is automatic in nature or has been automatized through practice or experience, then the task will impose low divided-attention demand. However, when traffic conditions become dense or when the driver has to negotiate complex manoeuvres in cluttered environments such as at busy intersections, then divided attention demands may exceed the attention capabilities of drivers. Spare attention capacity can be assessed by assigning drivers a secondary task as they perform different driving manoeuvres. McKnight & McKnight (1993) made a study of 150 subjects using simulated driving conditions in which individuals were expected to respond to different traffic situations by means of manipulation of simulated vehicle controls. The drivers were subjected to a variety of distractions such as: placing a cellular phone call, carrying a cellular phone conversation, carrying on an intense cellular phone conversation, tuning a radio and no distraction. It was found that there were significant differences in nonresponse with regard to age. In the age group above 50, there was an increase of about one-third in nonresponses under all of the cellular phone distractions. Similarly, Brown et al. (Brown et al., 1969 cited in Parasuraman & Nestor, 1993) found that more “effortful” activities such as judgement of distance in intersection crossing manoeuvres were more unfavourably affected than “automatic” activities such as gear changing, when drivers were assigned a reasoning task on a telephone headset. The literature does not report any studies that relate divided attention tasks to accident rate in drivers. With old age, efficiency in tasks that employ divided attention is known to show decrements especially when such tasks are of a complex nature; however, some studies do not corroborate this finding (McDowd & Craik, 1988; Salthouse, 1991 cited in Parasuraman & Nestor, 1993). Therefore, it is likely that in older drivers, divided attention tasks are able to gauge driving performance, but in younger drivers, they are not expected to make fruitful contributions to prediction of driving performance.

A state known as mild cognitive impairment (MCI) has been identified as a state that is some where between the state of cognitive changes of normal ageing and that of Alzheimer's disease (see Figure 2.3 & 2.4). The memory loss experienced in MCI is greater than that one expects in normal healthy ageing but still they do not meet the standard criteria for clinical diagnosis of Alzheimer's disease (Petersen et al., 2001). However, in MCI patients, the rate of progression towards clinically probable Alzheimer's disease is more accelerated than that of healthy individuals of the same age. The figure below shows the theoretical continuum for those subjects who progress from normal ageing to MCI and then to dementia. As can be seen in the figure, there is some overlap between the extremities between normal ageing and MCI and between MCI and between early dementia.

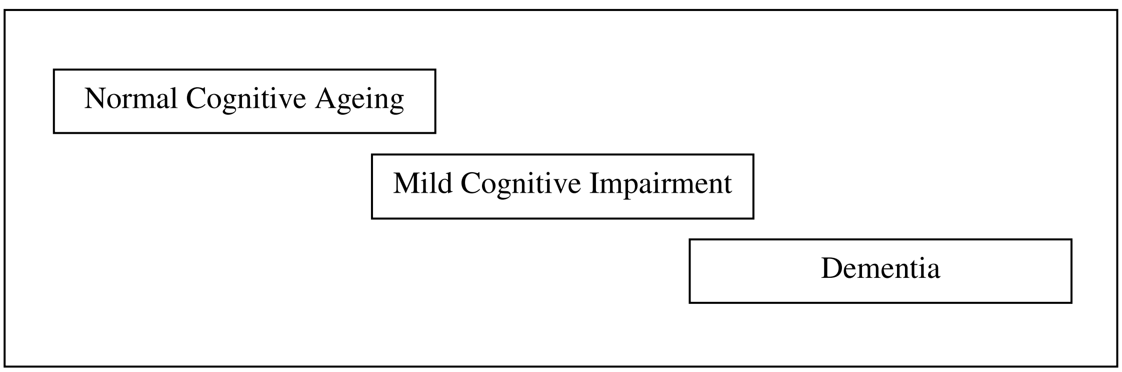



**Figure 2.3** Theoretical Continuum from normal cognitive ageing through to MCI and to dementia (Petersen, 2003).

As reported by Morris et al. (2001), MCI is diagnosed by the following observations: (a) proof of impairment in memory, (b) general cognitive and functional abilities are intact, and (c) diagnosis does not show dementia. Drivers afflicted with MCI also may exhibit reduced driving competence (Snellgrove, 2005). In normal cognitive ageing, there is general decrement in the speed with which information is processed, the efficiency with which new information is acquired also decreases, there is cognitive inflexibility and the working memory shows signs of reduction (Anstey et al., 2003; Rue, 1992; Nilsson, 2003; Salthouse & Meinz, 1995 cited in Snellgrove, 2005).

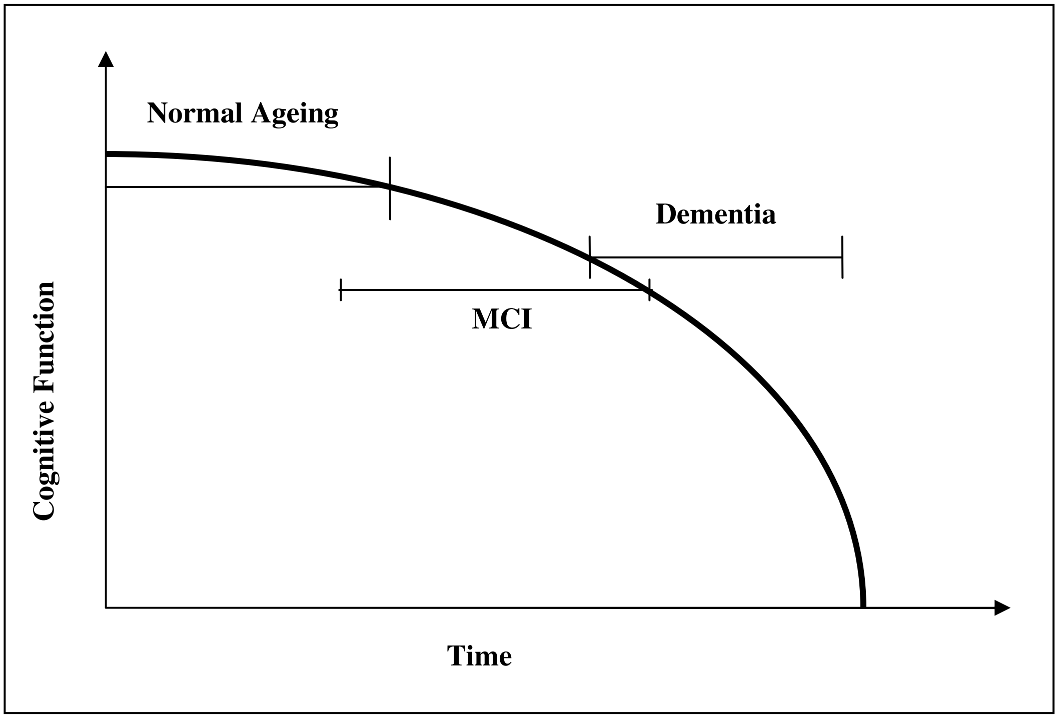



**Figure 2.4** Concept of Mild Cognitive Impairment (Huang, 2003)

Several scales are being used in classifying individuals along a continuum from normal ageing to the various stages of dementia. One such commonly used scale is the Clinical Dementia Rating Scale (CDR) (Morris, 1993 cited in Petersen et al., 2001).According to this scale, subjects classified under rating of CDR 0 are normal, CDR 0.5 have questionable dementia, CDR 1 have mild dementia, CDR 2 have moderate dementia and CDR 3 have severe dementia. Some analysts even regard CDR 0.5 as representing MCI, while other think that CDR 0.5 represents both MCI and mild Alzheimer's disease. A large percentage of subjects with MCI progress to dementia. Snellgrove (2005) reports that about 15 percent of individuals with MCI develop dementia within an year, 40 percent over 2 years, 53 percent in 3 years and almost 100 percent within 5 years. While as Petersen et al. (Petersen et al., 2001 cited in Lundberg, 2003) report that the conversion of MCI to AD (Alzheimer's disease) varies between 20 and 50 percent with an annual conversion rate from 10 to 15 percent. Petersen et al. (2001) report that the Mayo Alzheimer's Disease Research Center in Rochester, Minnesota made a longitudinal study of MCI subjects for more than 10 years and found that up to 80 percent converted to Alzheimer's disease status during approximately 6 years. Morris et al. (2001) and Bennett et al. (2002) (cited in Lundberg, 2003) report that progression of the condition depends on the extent of the impairment at baseline. According to the opinions of some researchers, MCI is not a separate disease, but rather it represents very early dementia (Fellgiebel, et al., 2004; Morris et al., 2001; Ritchie, et al., 2001 cited in Snellgrove, 2005), because of its high conversion rate to dementia. While Bozoki et al. (Bozoki et al, 2001 cited in Lundberg, 2003) have found out that MCI may sometimes be a stable condition and the risk of it being converted to dementia is elevated when indications of memory decline and impairments of other abilities are present in the same individual.

In early AD (Alzheimer's disease), patients are known to show impaired performance on memory tests, tests of delayed recall and tests framed to measure new learning (Petersen et al., 1994; Welsh, et al., 1991) and deficiencies of attention, executive function and language have also been noted, but it has been found that mild AD is better characterized by deficits in more than one cognitive domain (rather than by relying on memory alone) (Masur, et al., 1994). According to Petersen et al. (2001), neuropsychological tests alone cannot be used to diagnose Alzheimer's disease and the importance of clinical judgement in this context cannot be overemphasized. Snellgrove, 2005 (Davis & Rockwood, 2004; Kawas, 2003; Petersen, 2003 cited in Snellgrove, 2005) gives a list of the clinical criteria for MCI, which are shown below:

1. Memory Complaint, preferably corroborated by an informant.
2. Memory impairment at level of > 1.5 standard deviations of age/education psychometric norms.
3. Clinical Dementia Rating (CDR) score of 0.5.
4. Intact general cognitive function.
5. Essentially preserved activities of daily living (ADLs).
6. Not demented according to diagnostic criteria.

There are many factors that affect performance on neuropsychological tests, which include education, age, cultural background and illnesses other than Alzheimer's disease (Inouye, S.K., et al., 1993 cited in Petersen et al., 2001). Also, because of the commonalities of neuropsychological profiles in different types of dementia, neuropsychological tests cannot fully differentiate between the different types of dementia (Heyman et al., 1998; Ferman et al., 1999 cited in Petersen et al., 2001).

Greenwood et al. (1997) in a study of individuals with mild DAT (dementia of the Alzheimer type) and older subjects found out that there was a reduction in the control of spatial focus of attention in Alzheimer's disease subjects and to a lesser extent in advanced age. In a task that involved cued-visual search which allowed for cueing with variable precision of a target letter in an array of letters, the benefits of spatial cueing for the old-old group were greater than that of the DAT group but were lower than that for the young-old group. The authors finally concluded that the ability to control spatial attention lies on a continuum from healthy young adults to the young-old, through the old-old to individuals with mild DAT.

In another study, Parasuraman et al. (2000) used a cued visual search task to investigate the dynamic range over which spatial attention influences the identification of target by conducting visual search tasks to 42 participants that consisted of 15 non-demented "young-old" adults (65-74 years), 15 non-demented "old-old" adults (75-85 years) and 12 subjects having DAT (Dementia of the Alzheimer's type) in its early stage. The precues were valid, non-valid or neutral. In the non-demented subjects, it was found that as the precision of a spatial cue (that preceded the search target) increased (i.e. size of precued area decreased), the speed of detection (i.e. reaction time) of a target situated amongst distracters also increased. The cue size effect was greatly reduced in both the old-old and the DAT groups when compared to that of the young-old group, however the old-old and the DAT groups did differ in the range over which the effects were found; in the DAT group, spatial attention was confined to the cues that were the most precise. The authors explain that the partial similarity of results between the old-old and the DAT group may be due to the fact that there are some individuals in the old-old group that are in a very early preclinical stage of DAT. Finally, they conclude that the shrinkage in the dynamic range of spatial attention may reflect an underlying factor of the impairment in perceptual and memory functioning in early DAT.

## 2.7 Driving Simulator & On-road tests

Two kinds of errors are committed in the driving task: (a) low-frequency high-severity, and (b) high-frequency low-severity. If driving behaviour in diverse traffic environments/manoeuvres is objectively measured, it can disclose crucial and latent relationships between these two kinds of errors which can help in the prediction of driving safety in individual drivers (Wierwille et al., 2002 cited in Rizzo et al., 2005). Thus in order to accurately estimate the risk of an accident (which is a low-frequency high-severity error), it is necessary that a sufficient number measurable safety errors (which are high-frequency low-severity events) be evaluated/assessed (Rizzo et al., 2001).

Driving skills can be assessed either on a driving simulator or an on-road test. Because the driving simulator provides strictly controlled conditions in a synthetic environment, hence it provides a good opportunity to study the effects of cognitive impairment on driver errors. In a driving simulator, we can manipulate the task demand to put strain on a specific/ particular information processing stage in driving so as to generate a traffic scenario in which driver errors of one type or another are likely to be committed (Rizzo et al., 2001). A driving simulator can be programmed to provide information inflows to the driver that can not be achieved in the real world and then evaluate the specific driver reactions in a crash or hazardous scenario——an experimental procedure that is dangerous and unethical on the road. Simulators have the ability to replicate road conditions which incite drivers to make decisions in a safe environment. There are several advantages of using driving simulators vis-à-vis road tests or driving records in assessment of fitness to drive. Driving simulators provide the only way in which we can exactly reproduce experimental roadway conditions so that people can be tested under identical conditions and comparisons made. They also provide a safe environment to work in without the risks inherent in an actual road test and provide an opportunity to observe serious driver errors. Also, in a driving simulator, a driver can be subjected to standardised challenges that stress crucial cognitive abilities of driving. One drawback of simulation research involves simulator adaptation syndrome (SAS), which includes symptoms like nausea and sweating (Rizzo, 2004). SAS occurs due to a mismatch between visual cues of movement (which are enough) and inertial cues (which are scarce) and even occurs in simulators with a motion base. The severity of the impact of SAS can be reduced by using simulators that do not have large field of view displays or do not incorporate turns (especially left turns, right turns in Britain) so that peripheral visual field stimulation does not take place and movement cues do not sweep across the peripheral field. Also, the driving behaviour of drivers may be different in simulators since there is no danger of injury occurring compared to real-life driving situations. In a driving simulator, a multidimensional real-life task is translated to a two dimensional simplification; the congruence between the simulated and the real life tasks will occur to the extent to which they share common components (Rizzo et al., 1997). Even more important in this context is the level of psychological fidelity or functional equivalence of the simulator. However, advances in microelectronics, sensor, communication and control technology have provided for the more extensive automation of the evaluation of human-machine interactions (Rizzo et al., 1997) and resulted in machines that can more closely mimic reality. As reported by Rizzo (2004), driving simulators have been successfully used in assessing performance of drivers in conditions such as sleep apnea, drowsiness, alcohol and other drug effects , old age, Alzheimer's disease, Parkinson's disease or traumatic brain injury. In a study of simulated car crashes at intersections in drivers with mild to moderate Alzheimer's disease, Rizzo et al. (2001) tested drivers in a simulation that consisted of *"multiple "events" associated with potential crashes interspersed with uneventful highway segments."* In the study, 30 drivers drove a virtual road on the simulator and when they were within 3.6 seconds of an intersection, a cross vehicle made an illegal intrusion into the intersection. In order to avoid a collision with the cross vehicle, the driver of simulator had to perceive the intrusion and focus attention on the development and evaluate the situation. Next, an action plan had to be devised in order to deal with the hazardous situation by operating on the accelerator pedal, brake pedal or steering control. All reactions had to be performed under time pressure. Ideal response required that drivers release the accelerator, apply the brakes and make necessary steering corrections as required to remain within the traffic lane (i.e. safe avoidance). The authors found that subjects having AD (Alzheimer's disease) had a significantly increased risk of crash compared with non-demented drivers of similar age.

After examining a plot of steering wheel position, brake and accelerator pedals, vehicle speed and vehicle and lane position 5 seconds prior to the collision, the authors observed that in cases where collision did take place, inattention and inappropriate or slow control of response was exhibited by drivers. In some cases, drivers managed to avoid a collision with the intruding vehicle (the main hazard), but due to the late reaction/response, they experienced a secondary collision. It was noted that while driving on the uneventful section of the road prior to the intersection, no driver committed a safety error. Anderson et al. (2005) contend that since subjects having Alzheimer's disease or related disorders perform driving tasks adequately in familiar and ordinary traffic situations but cannot cope with unexpected or extraordinary circumstances, hence it is important that their responses to potentially dangerous situations be evaluated when assessing safety of such individuals. Fox et al. (Fox et al., 1998 cited in Schultheis et al., 2003) also argue that in order to predict driving adequately, drivers should be put through complex situations. In order to assess go/no-go decision-making by cognitively impaired drivers, Rizzo et al. (2003) (also cited in Rizzo, 2004) designed an abstract virtual environment (using soft ware) that consisted of a straight, flat, two lane road intersected by 100 crossroads. The task used a personal computer and the environment was displayed on a 21-inch video monitor. Tests were conducted on 16 subjects with neurological impairments and 16 neurologically normal subjects. The individuals drove and encountered a series of intersecting roads with gates that opened and closed as drivers approached. Drivers were directed to drive though the intersection quickly without hitting the gates with traffic signals correctly predicting imminent closure of gates at 80 percent of the gates. Inputs from steering wheel and accelerator /brake hardware peripherals was recorded. Measures included completion time, number of crashes into closed gates, number of stops at open gates, and number of successes (i.e., stopping at closed gates and going at open gates). More errors were committed by neurologically impaired drivers at gates and they took more time to complete the task. Also, in the study, the authors observed that *"……cognitively impaired drivers who had crashes at the gates or took longer to get through the task continued to show good control of the vehicle and did not exceed the lane boundaries, indicating that visuomotor control can be intact in drivers with decision making impairments, and that measures of visuomotor control in the driving task (such as* *steering and lane position variability) alone are not sensitive predictors of critical incidents caused by decision-making-impaired drivers."* Owing to the small field display, the subjects did not complain about simulator adaptation syndrome (SAS). Simulator output that may provide clues to measures of performance include steering wheel position (in radians or degrees), normalized accelerator and brake pedal position ( i.e. scale of pedal depression from 0 % -100 %), lateral and longitudinal acceleration (in terms of gravity), distance headway (meters), time to collision (seconds) and speed (km/h). Close calls (or near misses) can also be analysed to provide some kind of index of driver safety and if a sufficient number of observations are obtained, it may even be possible through the assessment of measurable safety errors to reliably reckon relative crash risk, even if no crashes take place (Rizzo et al.,1997). In a study of impaired response implementation for older drivers with cognitive decline, Rizzo et al. (2005) tested 48 drivers (mean age 73.5 years) with cognitive impairment caused by mild to moderate Alzheimer's disease and 101 (mean age 69.3 years) neurologically normal drivers in a driving simulator, in which they encountered a police car on the shoulder of the road. The primary purpose was to investigate situation awareness response to a roadway emergency in at-risk drivers. The reaction of drivers with cognitive impairment was more slow compared with neurologically normal drivers and they had more abrupt decelerations or failed to steer clear of the police car on the shoulder and the people situated near it. Even some impaired drivers stopped their vehicle in the middle of the road. The authors finally concluded that drivers with cognitive decline have a tendency to exhibit decreased situation awareness or poor executive control with regard to the selection and implementation of a potential evasive manoeuvre. Visual and neuropsychological measures of perception, attention, memory and executive function were able to predict the unsafe reactions of cognitively impaired drivers. Reger et al. (2004), while conducting meta-analysis of 27 primary studies to examine a relationship between neuropsychological functioning and driving ability for adults with dementia, found a significant relationship between neuropsychological functioning and driving ability as measured by on-road tests and non-road tests. However, the authors remarked that the higher correlations of non-road test groups with neuropsychological tests may be attributed to the tapping of purer skills deemed relevant to driving. The authors further added that although on-road tests may have higher ecological validity, some of the limitations of these tests include high cost of testing, subjectivity in scoring, inability to regulate control variables such as traffic flow, roadway conditions and other driver behaviour, which have the effect of decreasing the strength of relationship with neuropsychological tests. While on the other hand, non-road tests give the evaluators the ability to standardize traffic scenarios encountered by drivers by manipulating certain control variables e.g. such as traffic flows etc (variables that confound the results of road tests).

Using instrumented vehicles under actual road conditions, an on-road test can quantitatively access driving performance in the field. In a standard road test, the quantitative performance measurements that are made in an instrumented vehicle are free of human bias that tends to affect inter-rater reliability. Information on proximity, lane keeping, merging and following behaviour of the driver can be gathered through the aid of radar and video systems on board an instrumented vehicle. G-forces generated as a result of hazardous events like abrupt braking, swerving or loss of vehicle control can be detected with the aid of accelerometers (Rizzo, 2004). Specific criteria can be used to highlight instances where a critical incident may have taken place and used to filter the raw data generated from an instrumented vehicle. For instance, as drivers apply brakes or swerve to avoid hitting an obstacle, the longitudinal and lateral accelerometers measure g forces which highlight critical driving scenarios in the data stream (e.g. a value of 0.4g can indicate abrupt deceleration). Uc et al. (2005) utilized an instrumented vehicle to study driver identification of landmarks and traffic signs after a stroke. Thirty two drivers with stroke and 137 neurologically normal older adults during an experimental drive along a segment were asked to record sightings of specific landmarks and traffic signs. It was found that drivers having suffered a stroke distinguished significantly fewer landmarks and traffic signs and tended to make more at-fault safety errors compared with normal subjects. It was also noted that basic vehicular control as measured by standard deviation of steering wheel position (degrees), number of large (>6 degrees) changes in steering wheel position per minute and standard deviation of mean speed, on a straight segment (with no task load) was the same between the two groups.

Although, in a driving simulator, the reward and penalty structure is different from that of real life but likewise, in an on-road test, the natural pattern of driving is disrupted as the driver drives under the critical eye of the instructor or in someone else's vehicle (e.g. a test car equipped with dual controls). In an on-road test, the driver is exposed to traffic scenarios that he may endeavour to avoid in real life (Rizzo et al., 1997). Brown & Ott (2004) point out that in the presence of instructors during on-road tests, many drivers may be more anxious or more careful, which may not reflect their natural driving behaviour. Also, during on-road testing, if the driver is about to commit a certain error, the instructor will intervene in order to maintain safety and so the instructor is not able to know how the examinee might have handled that specific error. According to Rizzo (2004), state road tests were devised to find out whether novice drivers know the rules of the road and can apply them in a traffic environment; hence they are not an ideal tool for predicting accident involvement in experienced /skilled drivers who may have become impaired with time. Also, the author points out that to rely on real-life crashes as an assessment paradigm for driving skill is not credible as accidents are sporadic uncontrolled events that are difficult to objectively evaluate. In the assessment of driving performance on road tests, increase in variability of road test scores results from fluctuations in traffic and roadway conditions and subjective scoring by instructors which leads to a reduction in the strength of relationship with neuropsychological test scores (Rizzo et al., 2005).The high subjectivity in rating driving performance primarily results because not all evaluators/instructors use a structured checklist or other appropriate method of quantifying driving performance (Schultheis et al., 2003). Rizzo et al.(1997) suggest that it is reasonably expected of a licensed driver that when he is confronted with a hazardous situation (such as a short stop or an illegal intersection intrusion by another vehicle, which can be avoided/negotiated safely in most cases by most drivers), he should be able to recognize it and negotiate it safely. The simulator provides such an opportunity to test a driver. According to Rizzo et al. (2001), to predict driving performance through the observation of meaningful safety errors, it is necessary that the experimental driving task exposes the driver to a sufficient challenge. And according to the authors, this is one of the primary reasons why road tests which are tailored to provide driving routines that minimize threat of injury, are unsuccessful in predicting crashes in older cognitively impaired drivers (although they have considerable experience) who are at a higher risk of being involved in real world accidents. Lundberg et al. (1997) also highlight that with regard to the content of a standard driving test (road test), for experienced drivers with suspected cognitive impairment, the test does not pose a challenge big enough to bring out existing deficits in driving on account of the automatized nature of the driving task. McKnight & McKnight (1999) in a study regarding age related driver ability and performance deficits, administered measures of basic abilities and driving performance to a sample of 407 drivers over the age of 62 that comprised of drivers that had been referred to licensing agencies because of involvement in unsafe driving and a group of volunteers that had not been involved in traffic incidents. The authors concluded that since in an on-road test, practical control over site and evaluator differences in hardly realizable, therefore the road test is not an ideal tool for the detection of decline in driving abilities that occur as a result of ageing.

According to Bieliauskas (2005), the predictive power of neuropsychological tests can significantly improve if when assessing driving performance, the subjects (young & old) are made to face a challenging driving scenario (in a driving simulator) such as: (a) a wild animal all of a sudden runs into the road, (b) a vehicle intrudes from a side road/cross road/drive way from the right (left in Britain), (c) the vehicle in front of the driver suddenly stops. Freund et al. (2002) did a study to evaluate driving performance of cognitively impaired and healthy older adults in a pilot study that consisted of a small sample and compared on-road testing and driving simulation. In the on-road test and the driving simulator, the performance measures used were hazardous or potentially catastrophic errors, traffic violations and rule violations. There was significant correlation between average scores on the on-road test and the driving simulator (Pearson correlation coefficient of -0.670). Besides, a strong correlation was also found between hazardous mistakes (correlation coefficient -0.830) and lethal mistakes (correlation coefficient -0.816) and failing the road test. The authors finally concluded that simulated driving and on-road driving exhibited a strong mutual correlation for both cognitively impaired drivers as well as healthy old adults.

There is a possibility that drivers in a driving simulator may behave differently as there is no risk of injury compared to the driving situations that occur in real life, where there is a risk to life, limb and licensure are at stake. However, since the driving environment in a driving simulator has ecological validity (i.e. the driving maneuvers/hazards closely emulate the conditions that occur in real life and the same cognitive processes/decision making processes are at work in negotiating traffic scenarios and the simulated environment is supplemented by special audio effects that include engine and road noise) which can motivate drivers favourably. This is manifested by some recent validation studies that have demonstrated successful application of simulators (Fisher et al., 2002; George, 2003; Lee et al., 2002b) where deficits in performance on the simulator have been shown to correspond to shortfalls in on-road driving. Stress effects in simulators are minimum, however certain drivers using the simulator for the first time may experience some stress, which can be ameliorated through practice before taking the real test.

As reported by Rizzo (2004), a controlled auditory verbal (e.g. holding a conversation, performing mental arithmetic or using modern in-vehicle telematics such as cell phones and navigation devices) processing load (such as the Paced Auditory Serial-Addition Task [PASAT]) can be given to subjects (while driving) in order to scale the relevant impact / interactions in ageing and brain injury. However, in another study, Rizzo et al. (2004) assessed the effect of PASAT (while driving an instrumented vehicle) on 78 neurologically normal (mean age 71 years) and 82 subjects (mean age 75 years) who had impairments of selective attention but did not have diagnosable neurological disease. It was observed that imposing PASAT while driving resulted in reduced speed / steering control and greater number of at-fault safety errors were committed by the older drivers compared to the condition when PASAT was not imposed. But it was noted that there was no significant difference in the driving performance between the neurologically normal group and the attention-impaired group in both instances i.e when PASAT was imposed and base-line condition (PASAT not imposed). On reason for this could be that the PASAT is less emotionally engaging than for example talking to a friend on a phone as such an operation arouses more interest/attention in the distraction task. Also, the effects of distracter tasks (like PASAT / cell phones etc) on driving may be more pronounced if the driving conditions are more attention-demanding like driving through dense traffic, making complex manoeuvres and negotiating intersections etc. In this context, relevant to driving simulators, the authors (Rizzo et al.,2004) state:

*"Whether the safety tradeoffs that a distracted driver is willing to make in a driving simulator sufficiently resemble those that a distracted driver would execute while driving a real car on a real road………"*

## 2.8 Summary

About 40 percent of the driving population will be over the age of 60 by the year 2020 in the UK and currently, several hundred thousand drivers with dementia hold driving licenses. It is well known that the number of motor vehicle crashes per unit distance of automobile travel is "U"-shaped, with risk increasing slightly between the ages of 55 and 60, but greater increases in risk with each successive five-year interval. The crashes involving older drivers more often occur in complex situations, where the driving task is not self-paced and there is a particular risk of cognitive overload. A number of driver error categories have been designated as playing a role in the accident causation of older drivers. These include failures of perception, attention, cognition and action.

In the crashes of older drivers, visual and cognitive decrements are the most important causal factors. Cerebral visual impairments in drivers makes them liable to *"look but not see"*, despite the fact that information load is low. In old age, automatic routines remain relatively well preserved but older people find it very difficult to inhibit automatic processes in suddenly changing (and unexpected) situations.

Alzheimer disease is more prevalent in the old age group. Decline in cognitive functioning is accompanied by concomitant decline in driving ability. For individuals that have moderate to severe dementia, there is strong consensus that they should not drive, however decisions regarding those having mild dementia are problematic. The most challenging assessment and decision for the physician / licensing authority as regards fitness to drive lies in drivers who are questionably demented or are in a state of very mild dementia. The prerequisite for driving cessation should be impaired competence for driving rather than a diagnosis of dementia/disease condition. Individuals with mild dementia should be considered for specialized assessment with regard to driving competence. Cognitive deficits in Alzheimer's disease and other dementias have serious implications for driving, since drivers are not able to tackle multiple stimuli simultaneously (show decrements in divided attention and selective attention abilities), maintain sustained attention, quickly respond to hazardous situations, judge distances, correctly interpret traffic control devices and exhibit a propensity for driving slowly relative to the speed limit. With regard to neuropsychological tests that facilitate in determining whether a dementia patient is safe to drive, numerous neuropsychological tests are not sensitive and specific enough to tap the behavioural and cognitive features /domains that may affect driving. Also, many psychometric tests (cognitive tests) are not process-specific, but are rather reflective of general processes and hence these tests may tap general cognitive status, rather than skills necessary for driving.

Driving skills can be assessed either on a driving simulator or an on-road test. A driving simulator can be programmed to provide information inflows to the driver that can not be achieved in the real world and then evaluate the specific driver reactions in a crash or hazardous scenario——an experimental procedure that is dangerous and unethical on the road. Driving simulators provide the only way in which experimental roadway conditions can be exactly reproduced so that people can be tested under identical conditions and comparisons made. One drawback of driving simulators is that a small proportion of drivers experience simulation sickness syndrome. Nausea, disorientation and ocular problems such as eyestrain, blurred vision and eye fatigue have been reported as some of the indicators of simulation sickness in fixed-base simulators.

Subjects having Alzheimer's disease or related disorders perform driving tasks adequately in familiar and ordinary traffic situations but cannot cope with unexpected or extraordinary circumstances, hence it is important that their responses to potentially dangerous situations be evaluated when assessing safety of such individuals; this can only safely be evaluated in driving simulators. Although on-road tests may have higher ecological validity, some of the limitations of these tests include high cost of testing, subjectivity in scoring, inability to regulate control variables such as traffic flow, roadway conditions and other driver behaviour, which have the effect of decreasing the strength of relationship with neuropsychological tests. In a driving simulator, the reward and penalty structure is different from that of real life but likewise, in an on-road test, the natural pattern of driving is disrupted as the driver drives under the critical eye of the instructor or in someone else's vehicle (e.g. a test car equipped with dual controls). In the presence of instructors during on-road tests, many drivers may be more anxious or more careful, which may not reflect their natural driving behaviour. Also, during on-road testing, if the driver is about to commit a certain error, the instructor will intervene in order to maintain safety and so the instructor is not able to know how the examinee might have handled that specific error.

# 3 Neuropsychological Tests

## 3.1 General

The objectives for this chapter were to identify the cognitive attributes that are relevant / crucial for the driving task and then based upon the identified cognitive constructs come up with a battery of cognitive tests along with their description and detailed testing/scoring procedure. The key functions required for driving are vision, cognition and motor functions. This study will deal primarily with cognitive functions, although, there is bound to be considerable overlap between functions. The driving environment is quite complex, where vehicle control has to be negotiated in a cluttered environment through the simultaneous use of both central and peripheral vision to process both primary (high priority) and secondary (low priority) visual tasks and where the prediction of important events in time and space is difficult/uncertain. According to Hartman (Hartman, 1970 cited in Kito et al., 1989), visual information constitutes over 90 percent of the stimulus input in driving. Specifically, cognitive functions that support adequate visual processing and the coordination and integration of perception and motor skills are very essential for safe driving. Multiple cognitive domains are called upon when an individual is negotiating driving scenarios/situations. We must bear in mind that many of the cognitive constructs themselves are interrelated (Ball et al., 2004) and interact with each other (Anderson et al., 2005). According to Rizzo (2004), the coordination of several ongoing processes including attention, perception, memory (declarative, procedural, and working), and executive functions (decision making and implementation) is required for safe driving.

Reger et al.(2004) conducted meta-analysis of 27 primary studies to examine the relationship between neuropsychological functioning and driving ability for adults with dementia and concluded that visuospatial skills and attention were most helpful in screening at risk drivers. In a study by Shinar (Shinar, 1978 cited by Parasuraman & Nestor, 1991), it was estimated that 25 to 50 percent of motor vehicle crashes are the result of driver inattention. Rizzo et al. (2005) in their study found that visual and neuropsychological measures of perception, attention, memory and executive function were able to predict the unsafe reactions of cognitively impaired drivers. Lundqvist et al. (1997) put 29 brain damaged patients and controls to neuropsychological tests, simulator driving and on-road tests. Based on the outcome of their findings, they recommended using neuropsychological tests that measure attention, information processing and executive functions. Christie et al. (2001a) evaluated 39 subjects with head injury or generalized brain damage by giving them neuropsychological and road tests. Logistic regression was used to identify 5 tests that were significant predictors of driving skill. These tests measured visual memory, executive abilities, spatial awareness and attention. Van Zomeren et al. (1988) reported that controlled attentional processes are impaired after head injury, while automatic processes are not. Nouri et al. (1987), while investigating the relationship between cognitive ability and driving after stroke, in which 39 pre-stroke drivers were assessed using a battery of cognitive and on-road tests, finally concluded that cognitive tests which involve complex reasoning skills appear most highly related to driving performance.

Duchek et al. (1998) examined the relationship between visual attention measures and driving performance in healthy older adults and individuals with mild dementia of the Alzheimer type. Subjects were also assessed for on-road driving. They finally concluded that selective attention (i.e. visual search performance) is related to driving skills in the early stages of DAT (Dementia of the Alzheimer type). Parasuraman et al. (1992) evaluated subjects with mild to moderate dementia and concluded that patients in the early stage of DAT show impairments in the ability to switch or disengage attention, while as they retain the ability to focus attention to a spatial location. Parasuraman & Nestor (1993) reported that the ability to switch attention is also correlated with driving performance in normal individuals.

During the course of normal ageing, some cognitive functions e.g. spatial orientation and perceptual speed have been found to decline (Schaie, 1996). Also, older adults have problems related with the spatial and temporal integration of items into larger units, which may reflect their deficits in working memory (Welford, 1985 cited in Lundberg, 2003). Working memory declines with age (Foos & Wright, 1992; Salthouse, 1992 cited in Ball et al., 2004). Working memory can be described as the ability to maintain or manipulate information while another task is performed concurrently. Visual search is an important element of driving and is crucial for detecting potential road hazards. Studies have shown that when cognitive processes including working memory are loaded, the efficiency of visual search is affected (Wood et al., 2006). The "frontal ageing hypothesis" stipulates that in the ageing process, the prefrontal cortex is one of the most affected areas of the brain (Bieliauskas, 2005). Since the prefrontal cortex has an important contribution in exercising executive control (Funahashi, 2001), putting more emphasis on executive functioning is in harmony with this hypothesis. In normal cognitive ageing, there is a general decrement in the speed with which information is processed, the efficiency with which new information is acquired also decreases, there is cognitive inflexibility and the working memory shows signs of reduction (Anstey et al., 2003; Rue, 1992; Nilsson, 2003; Salthouse & Meinz, 1995 cited in Snellgrove, 2005). Greenwood et al. (1997) in a study of individuals with mild DAT (dementia of the Alzheimer type) and older subjects found out that there was a reduction in the control of spatial focus of attention in Alzheimer's disease subjects and to a lesser extent in advanced age. The authors finally concluded that the ability to control spatial attention lies on a continuum from healthy young adults to the young-old, through the old-old to individuals with mild DAT.

Visuospatial skills give an individual the ability to determine where things lie in relation to each other and to himself/herself. Visuospatial skills play their role in a number of tasks relevant to driving that include: (a) the correct positioning of the vehicle in the roadway environment (b) spatial manoeuvring on the roadway (c) the judgement of distances to other objects or vehicles (e) monitoring and forecasting the spatial positioning of other vehicles. Divided attention enables a driver to simultaneously monitor two or more stimulus sources. The driver has to allocate his attention simultaneously to various sub-tasks which include steering (and maintaining a desired speed), navigation, obeying regulations and warnings, manipulating in-vehicle mechanical and environmental systems (including entertainment systems), communicating, monitoring events outside and to a lesser extent inside the vehicle and keep a watch over the instrumentation panel. For example, the driver has to monitor the environment and at the same time, has to maintain lane position. In complex and cluttered driving environments selective attention skills come into play as the driver has to focus and shift attention between different elements /features/ developments/changes in the driving scene and then has to preferentially process particular events from within the scene so that an appropriate response can be formulated and executed. According to the British Psychological Society (2001), Executive functions " govern ability to anticipate, plan ahead, make decisions, self-monitor, and change a plan of action, sometimes instantaneously ". Executive functions enable drivers to organize information and plan complex behaviour. As a result drivers who have executive dysfunction will not be able to develop appropriate strategies (which might involve complex tasks) and programme their movements when they face a novel or distracting cue. Executive functions also make use of working memory capacity. Executive skills are crucial to monitoring the roadway environment and adapting to the traffic at hand. This adaptation can be brought about by carrying out vehicle manoeuvres and responding in an appropriate manner to other vehicles or road users. One example when executive functions are brought into play is in the selection and implementation of an evasive manoeuvre performed by a driver who is travelling above a certain speed threshold. If all of a sudden a pedestrian steps onto the road, the driver realizes that the sight distance available is smaller than the perception reaction and stopping distance and so if he or she brakes, the vehicle would still collide with the pedestrian. Therefore he/she makes a decision to swerve to avoid hitting the pedestrian. But then he has two options: (a) swerve to the adjacent lane (b) swerve to the shoulder. If he/she swerves to the adjacent lane and if traffic is present in these locations it can result in a near miss or a secondary collision. Swerving onto the shoulder would be a better option if there is no other obstacle present on the shoulder. It is here that we find that processing speed (cognitive processing speed) also has a lot of bearing on the outcome of the event. The speed of information processing is also crucial to driving because it is a moving environment. One draw back of the simulator compared to real world driving is that in the real world, the pedestrian may take countermeasures to avoid an accident (i.e., he may step back, change direction or increase walking speed), whereas in the simulator, once a pedestrian steps onto the roadway, he will continue with his course i.e., he is not intelligent. The importance of Executive function for safe driving is also highlighted by Freund et al. (2005a). According to the authors, the automatized and procedural skills which are learned over a long period through driving experience do not protect the older driver from making errors in the presence of executive dysfunction. Six neuropsychological tests have been identified to cover key cognitive domains necessary for safe driving. These tests also assess a broad range of cognitive skills and are in the public domain with the exception of one test (i.e. UFOV test). Significant diversity with regard to administration of these tests is exhibited as they include paper and pencil tests, listening test and visual computer controlled test. Also, these tests are quite sensitive to the effects of ageing and to a range of diseases that are well known to impair driving performance. The chosen tests possess good reliability i.e., the measure of the test should be stable and consistent if it is to be used to gauge "an enduring traitlike characteristic such as the quantity of one's processing resources" (Salthouse et al., 1989). In other words, the characteristic which is being measured by the test should remain stable over time. Although according to Salthouse & Kausler (Salthouse & Kausler, 1985 cited in Salthouse et al., 1988), at least a moderate reliability (e.g., $r > 0.6$) is required. With regard to neuropsychological tests we must bear in mind that a single test does not reflect a pure measure of a single cognitive domain, but rather each test taps more than one cognitive domain. Also, each test only partially taps a specific domain. Therefore, keeping these key points in view, we have selected more than one test that taps the same domain that is very critical relevant to the driving task e.g., there is more than one test that taps visuospatial abilities and attention (because these are highly crucial domains relevant to driving). Tests that measure vocabulary and general information have not been included as they are more dependent on knowledge that has been acquired previously and do not gauge current processing efficiency of individuals; also, there is a positive correlation between age and opportunity to acquire knowledge (Salthouse et al., 1988).

## 3.2 Trail Making-B Test

The trail Making-A and Trail Making-B are both timed paper-and-pencil tests. The original constructions of these tests took place in 1938 and were adopted by the US Army as a part of Army Individual Test Battery in 1944 (Strauss, et al., 2006). This test takes about 5 to 10 minutes to administer (Strauss, et al., 2006). In both tests, 25 circles are randomly distributed on a sheet of white paper that is 8 ×11 ½ -in in size. In Trail Making-A, the distribution of the numbers is from 1 to 25. The measure of the test is the time that it takes for a subject to draw a continuous line correctly connecting the circles in numerical order. In Trail Making-B, either there is a number (from 1 to 13) or a letter (from A to L) inscribed inside each circle. The measure of the test is the time it takes for a subject to draw a continuous line correctly connecting the circles in numerical and alphabetical order alternately (i.e. from 1 to A to 2 to B, etc). Both tests require attention and visual scanning skills. However, on account of its simple structure, Trail Making-A test assesses simple cognitive processing speed, whereas Trail Making-B mental flexibility and complex cognitive processing speed (Groth-Marnat, 2000). Lezak et al. (Shum et al., 1990 cited in Lezak et al., 2004) report that the motor component of this complex visual scanning test is such that motor speed and agility of the examinee strongly contributes to success on the test. Demands on visual search, speed of processing, and divided attention (in case of Trail Making-B) are placed by these tests (Rizzo et al., 2000).

Although both tests require non-verbal monitoring, Trail Making-B is comparatively cognitively more demanding as it requires switching of attention between numbers and letters (McKenna, 1998). Ballard et al.(Ballard et al., 1993) have described Trail Making-A as a test that measures visuospatial scanning abilities and motor sequencing skills, whilst Trail making-B requires some abilities over and above those required for trail Making-B, especially greater visual search skills. They report that these tests are sensitive to age and were 81 percent effective in diagnosing brain damaged subjects. However Trail Making-B showed a higher correlation with driving performance (i.e. -0.42 ), which was significant compared to Trail Making-A. Stolwyk et al. (Stolwyk et al., 2006) also failed to find a correlation between Trail making-A and most driving measures. However, trail Making-B correlated strongly with a wide range of driving performance measures in participants with Parkinson's Disease. Although, Trail Making-B has the same psychomotor requirements as that of Trail Making-A, but it has significantly increased Working Memory demands, because information has to be concurrently manipulated. As pointed out by Crawford et al. (Crawford et al., 1992), while taking Trail Making-B test, due to the double task of keeping track of alphabets and numbers in the mind, a person's working memory is loaded.

Wang et al.(Wang et al., 2003) in their report on assessing and counselling older drivers have also highlighted the fact that association between poor performance on Trail-Making-B and poor driving performance, has been demonstrated by numerous studies. They further underline that this test specifically assesses working memory, visual processing, visuospatial skills, selective and divided attention and psychomotor coordination. The divided attention aspect of the test is due to the fact that participants have to keep track of numbers and alphabets while at the same time have to search for other target circles. The Trail Making-B test has also shown high sensitivity to the progressive cognitive declines that take place during the course of Dementia (Storandt et al., 1984 cited in Stutts et. al., 1998). As reported by Carr et al. (Reger et al., 2004 cited in Carr et al., 2006), in older adults with dementia, poor driving outcomes (using driving simulators and road tests) has been correlated with poor performance on the Trail Making tests; and that the test relies on memory, visuospatial skills, attention and executive function. Perry & Hodges (Grady et al., 1988 cited in Perry & Hodges,1999) report one of the few longitudinal studies conducted on Alzheimer's disease patients, who were also given the Trial Making-B test; deficits on the test were evident after impairment in episodic memory but before visuospatial and language dysfunction could take effect. They report that the test also taps many aspects of executive function.

In a study in Maryland (Ball et al., 2006), in which 1910 older drivers aged 55 to 96 were evaluated in Motor Vehicle Administration (MVA) sites, it was found out that subjects taking 147 seconds or more to complete the Trail Making-B Test (90th percentile) were 2.01 times more likely to be involved in a motor vehicle collision as the subjects who completed the test in less time. Performance declines on Trail Making-A and Trail Making-B with advancing age (Strauss et al., 2006). Trail Making Test-A and Trail Making-B tests have exhibited interrater reliability coefficients of 0.94 and 0.90 respectively (Fals-Stewart, 1991 cited in Strauss, et al., 2006). Test-Retest Reliability for Trail Making-B is adequate for the most part (Strauss et al., 2006). As reported by Strauss et al. (Strauss, et al., 2006), an adequate coefficient (0.79) and high coefficient (0.89) were obtained for Trail Making-A and Trail Making-B respectively, by Dikmen et al.

(Dikmen et al., 1999) after examining 384 normal / neurologically stable adults aged 15 to 83 years; these participants were retested 11 months after the initial testing session. According to Reitan (Reitan, 1958), good performance on the test requires that the subjects be alert and have concentrated attention on the task. Overall Trail Making-B test is a complex measure of multiple cognitive functions which also includes the use of greater visual search skills (visual scanning and tracking) compared to Trail Making-A. Efficient visual search especially in visually complex driving environments (encountered in cities and urban areas) for potential hazards is an indispensable element of the driving task. Another example could be when a driver has to search the environment for a specific target in a relatively complex surrounding where many distracters are present (e.g., a street sign at an intersection). Also, in Trail Making-B, the examinee deals with more than one stimulus (thought) at a time, shows cognitive flexibility (opposite of cognitive rigidity) in shifting the course of an ongoing activity and has working memory demands, therefore it is more relevant to the driving task. Cognitive rigidity can create severe problems for a driver if he encounters a suddenly changing situation (e.g., like intrusion of a cross vehicle in an intersection). To exploit the element of parsimony, we opted for the Trail Making-B (compared with using both Trail Making-A & -B) test.

### 3.2.1 Trail Making-B Test Procedure and Scoring

In Trail Making-B Test, 25 circles are randomly distributed on a sheet of white paper that is 8 ×11 ½ -in in size; either there is a number (from 1 to 13) or a letter (from A to L) inscribed inside each circle. The measure of the test is the time it takes for a subject to draw a continuous line (without lifting the pencil from the paper) correctly connecting the circles in numerical and alphabetical order alternately (i.e. from 1 to A to 2 to B, etc) as quickly as possible. Errors (sequencing errors), if committed are immediately pointed out by the examiner, who instructs the examinee to continue with the rest of the test from the last correct connection. The stopwatch does not stop during error correction. Therefore, errors contribute to the extent that additional time is needed for corrections. Poorer performance on the test is reflected in longer completion times. Strauss et al.(Strauss et al., 2006) report that a number of authors include a time limit constraint of 5 minutes

(300 seconds) in order to reduce testing time and frustration. Staplin et al. (Staplin et al., 2003) in their study regarding the Maryland Pilot Older Driver Study used a maximum time limit of 6 minutes (360 seconds), at which point the test was discontinued. However, Strauss et al.(Strauss et al., 2006) report the calculation of time scores on prorata basis when the test was not completed within the time limit. In that case, the time score (in seconds) was derived by prorating by dividing the number of seconds allowed for the test by the number of circles completed, and then multiplying this "time per circle" by 25 to get the test score in seconds. The test is preceded by a sample test to ensure that the examinee has understood the test procedure and task. However, the sample test is not the same size as the actual trail Making-B Test, only 4 numbers (i.e. 1 to 4) and 4 letters (A to D) are included for rapid orientation. The practise session is not timed. Detailed Test Procedure is in Appendix A.

## 3.3 Clock Drawing Test

The clock drawing Test is a brief test that is easily administered and is frequently used in dementia evaluations. Its introduction was brought about in the early 1900s as an indicator of constructional apraxia (McDowell, 2006). In this test, verbal instructions are given to a subject to draw a clock (analog clock), put in all the numbers and set a particular time. The test takes less than 5 minutes to administer. When a verbal command is given that a clock should be drawn, to comprehend the instructions, the examinee must possess adequate language skills. The visuospatial features of the clock must have a representation for him as well as a mechanism by which this representation is recalled. Then, in order to translate the mental representation into a motor program for drawing, visuoperceptual and visuomotor processes are needed. The spatial layout of the component features of the clock is guided by visual perceptions. The accurate representation of features on both sides of the clock (i.e on the left and right side of the letter 12) is ensured by hemi-attentional processes. Motor output is monitored by visual perceptions and corrections are incorporated through control processes of executive functions. The graphomotor representation of the digits/numbers is brought about by the linguistic system as an output (Freedman et al., 1994).

Therefore, although Clock Drawing seems a simple task but this free hand procedure depends upon multiple cognitive functions organized in diverse cerebral regions (Groth-Marnat, G. (2000). According to Groth-Marnat (2000), some of these functions are auditory language skills, memory (storing instructions for a certain time setting), visuospatial ability, perceptual-motor facility, linguistic skills (for drawing numbers) and executive functions for planning and organization. Sunderland et al. (Sunderland et al., 1989 cited in Ferrucci et al., 1996) considers it a simple test focusing on visuospatial and constructional abilities. McDowell (McDowell, 2006) reports that recently it has been shown that disturbance of visuospatial skills are an early sign of dementia. Deficits in visuospatial abilities and abstract thinking are revealed and characterized by distortions in the placement of numbers on the clock face and hands to designate specified times. Its importance is reinforced by the fact that when some persons who are in the initial stage of dementia are administered verbal tests, they show fairly good performance; but when tests of visuospatial and construction abilities are administered, they fail dramatically (Moore & Wyke, 1984 cited in Ferrucci et al., 1996).

Strauss et al.(Strauss et al., 2006) report that for the screening of dementia, the Clock Drawing Test is frequently recommended. However, it should not be used as a stand alone test for dementia but rather as a supplement to other methods or used in conjunction with other tools (McDowell, 2006; Tuokko et al., 2000). If it is compared to other dementia scales, most have primarily verbal content, while as the Clock Drawing Test is dependent on visuospatial, constructional, and higher-order cognitive abilities which includes executive aspects as well. Alzheimer's Disease and functional disabilities are associated with impaired executive control functions, and it has been found that the Clock Drawing Test correlates with Executive Control functions (Royall et al., 1998 cited in Freund et al., 2005a) and also correlates with disease progression (Sunderland et al., 1989 cited in Freund et al., 2005a). It has been revealed by cross-sectional studies that age has an effect on performance on the Clock Drawing Test and this decline in performance is exhibited particularly after the age of 70 years (Freedman et al., 1994; Strauss et al., 2006). Incorrect representation of the proportion of the hands and placing the minute hand incorrectly are among the most common errors (Freedman et al., 1994). According to Strauss et al. (Strauss et al., 2006), the ability to position the hands bears importance for the detection of dementia. IQ, education and ethnicity have a low influence over the Clock Drawing Test (Shulman et al.,1993).

The higher-order cognitive skill of executive functioning required for the clock drawing test in the form of planning and organization is a critical component of safe driving because driving requires the integration of behaviour and instantaneous and appropriate response has to be made to changing situations in the traffic scene. Besides, the test is quick and easy to administer and is non-threatening (i.e., is well tolerated by examinees and examiners). Variations exist in the manner in which the Clock Drawing test is administered and scored. A number of these methods are discussed in the following paragraphs with their merits and demerits. Of the methods discussed, the Freedman method (Freedman et al., 1994) was selected.

### 3.3.1 Time specification for Clock Drawing Test

Shulman (Shulman, 2000) reports that various times have been used such as "3 o'clock" i.e. 3:00, "5 after 8" i.e. 8:05 and "45 after 2" i.e. 2:45 and "10 after 11" i.e. 11:10. According to Freedman et al.(Freedman et al., 1994), the most widely used time settings are: "10 after 11" i.e. 11:10, "20 after 8" i.e. 8:20 and "3 o'clock" i.e. 3:00. Kaplan (Kaplan, 1988) recommends using the "10 after 11" time setting in the instructions. This setting involves the placement of hands in both sides of the hemispace (i.e on the left and right side of the letter 12) and thus facilitates the identification of hemilateral or hemianopia (Strauss et al., 2006) and also has the advantage that it puts greater demands on the executive system, which is mediated by the frontal lobes of the brain. Patients with frontal lobe dysfunction are impaired in abstract thinking and as a result tend to make "stimulus-bound" errors wherein information is processed at a more perceptual rather than a semantic level (Freedman et al., 1994). When an examinee is instructed to set the time at "10 after 11", a recoding process is involved wherein the "10" has to be recoded and the minute hand set on the number "2". Since the number "10" on the clock is adjacent to the number "11", Freedman et al.(Shallice, 1982 cited in  Freedman et al., 1994) highlight that patients (subjects with frontal lobe impairement) who have a tendency to be "pulled" (frontal pull due to executive dysfunction) to the perceptual features of the command are likely to commit the stimulus-bound error by placing one hand (of the clock) on the "10" and the other on the 11. Alternatively, when an instruction for a time setting of "20 after 8" is given, the examinee has to recode the "20", because the clock face is devoid of the number "20" to "pull" the subject and therefore, other abnormal responses can occur such as the minute hand may be placed just after the eight or the minute hand may be set at the "2", because it is similar to the "20" (Freedman et al., 1994). In the "3 o'clock" time, the "o'clock" has to be recoded to "12" to represent correctly the minute hand. Since the time "10 after 11" appears to be the most sensitive to neurocognitive dysfunction, therefore we selected this time to be set by the examinees, although in the original study by Freedman et al. (1994) they used a time setting of 6:45.

### 3.3.2 Clock Drawing Test Scoring Methods

Variations exist in the manner in which the Clock Drawing test is administered and scored. Examinees are generally presented with a blank piece of paper or a predrawn circle 4 and 5/8 inches in diameter (without the numbers and hands etc) and asked to draw a clock and set a particular time; these methods are called the "free-drawn" and the "pre-drawn" methods respectively. The rationale behind using the "pre-drawn" method is that the clock drawing performance is focused on the numbers and hand placement and as a result, some difficulties that are inherent in procedures in which the examinee draws the circle as well are circumvented. Because if a circle is poorly drawn, it confounds the remainder of the clock drawing performance (Tuokko et al., 2000). For example, an examinee may not draw a circle large enough for the numbers to be put in, or may be asymmetrical which may affect the arrangement of numbers.

According to Freedman et al.(Freedman et al., 1994), the "free-drawn" method constitutes all of the elements of the clock drawing task. They argue that evaluation of the contour is crucial because if a person can not draw an acceptable contour, it can be considered pathological at any age. Also, in the instructions many examiners instruct the examinees to "draw the face of a clock, put in all the numbers, and set the time at ……" Other examiners do not give instructions for time setting at the beginning but only after the examinee has drawn the clock face and numbers. These examiners are of the view that if the time setting is known in advance, it may influence how the subject proceeds through the drawing. For example, if the time is known in advance and is “10 after 11”, an examinee may proceed by drawing “11,” “12,” “1,” and “2,” in that order, instead of starting with “12” and sequentially drawing the numbers clockwise. As a result of such untypical behaviour, errors in spatial organization may result (Freedman et al., 1994). In the instructions for the time setting, Shulman (Shulman, 2000) recommends that the instructor should not use the words “hand” in the instructions.

The differing scoring systems that are in place, place differing emphasis on visuospatial, executive, quantitative and qualitative issues (Kaplan, 1990 cited in Shulman, 2000). All have good interrater and retest reliability, and most are correlated with one another (Strauss et al., 2006). However the correlation is not so high as to make them interchangeable (Tuokko et al., 2000). Depending on the scoring method and sample composition, sensitivity (i.e., percent of individuals diagnosed with probable Alzheimer’s disease who score in the abnormal range of the test) and specificity (i.e., percent of control individuals who score in the normal range on the test) varies (Strauss et al., 2006). Some of these methods are discussed below:

<u>Mendez:</u> The Mendez scoring system (Mendez et al., 1992) is also called the Clock Drawing Interpretation Scale (CDIS) for testing on patients with dementia. The method is “free-drawn” and the time specified is 11:10. Better performance is reflected in higher scores. For scoring purposes, 20 individual items are evaluated and each item carries 1 point each and therefore, the best score is 20. This scoring system consists of: 3 items relevant to general characteristics of the clock (3 points), 12 items relevant to presentation and placement of number (12 points) and 5 items relevant to assessing the existence and placement of hands (5 points). A score of less than 18 suggests impairment. Sensitivity and specificity of the test varies from 88 to 94 percent and 26 to 65 percent respectively (Strauss et al., 2006). However, in this method, the full credit can still be obtained without the hands indicating correct time.

<u>Shulman:</u> The Shulman method of scoring (Shulman et al., 1986; Shulman et al., 1993; Shulman, 2000) uses the “pre-drawn” method and the time specified is 11:10; the total score is 5. The highest score is given to an intact clock and the clock that is impaired receives the lowest score. A “perfect” clock receives 5 points and a clock that has minor

visuospatial errors receives 4 points. If the visuospatial organization is well done but the time of 11:10 is inaccurately represented, then 3 points are awarded. If the visuospatial disorganization of numbers is moderate, such that it is impossible to accurately denote "10 after 11", the clock receives 2 points. A clock is awarded 1 point if there exists a severe level of visuospatial disorganization. If there is inability on part of the examinee to make any reasonable representation of a clock, a zero score is awarded. A score of less than 4 suggests impairment. Sensitivity and specificity of the test varies from 81 to 93 percent and 48 to 96 percent respectively (Strauss et al., 2006).

<u>Tuokko:</u> The Tuokko method (Tuokko et al., 1992; Tuokko et al., 1995) uses a "pre-drawn" clock and the time specified is 11:10. The test is scored by tallying the number of errors. In this method, scoring is achieved by using 25 different error types, which can be summed on seven subscales: omissions, perseverations, rotations, misplacements, distortions, substitutions, and additions. To assess the level of impairment, total error score can be obtained by adding the seven subscales. Low scores suggest minimal impairment and a score of zero indicates error-free performance. There is no ceiling for the total number of errors but in one study (Tuokko et al., 2000), a maximum score of 41 was obtained. A score of greater than 2 errors suggests impairment. To enhance the accuracy in judging the misplacement of numbers, a transparent overlay with marked zones of acceptable positions of numbers in used in this method. Shulman (Shulman, 2000) reports that normal elderly and Alzheimer's disease patients were significantly differentiated by four error categories (omissions, distortions, misplacements, and additions). Sensitivity and specificity of the test varies from 91 to 92 percent and 50 to 86 percent respectively (Strauss et al., 2006).

<u>Wolf-Klein:</u> The Wolf-Klein method (Wolf-Klein et al., 1989) uses a "pre-drawn" clock. It uses a rating scale that varies from 1 (most impaired) to 10 (normal). The Wolf-Klein method uses 10 hierarchical clock patterns that were developed from a pilot study of over 300 patients; this classification of ten categories of errors facilitate the scoring process. A score of less than 7 suggests impairment. Sensitivity and specificity of the test varies from 39 to 79 percent and 72 to 95 percent respectively (Strauss et al., 2006).

<u>Sunderland:</u> The Sunderland method (Sunderland et al., 1989) uses a "pre-drawn" clock and the time specified is 2:45. Six researchers/raters were asked by Sunderland et al.

(Sunderland et al., 1989) to categorize 150 completed drawings into 10 ordered categories that varied from the "best" representation of a clock to the "worst" representation of a clock. After analyzing these categories, descriptive criteria were developed for ranking clocks on a scale from 1 to 10 (higher numbers representing better performance). Examples of the criteria are: "5. Crowding of numbers at one end of the clock or reversal of numbers" and "7. Placement of hands is significantly off course". A score of less than 6 suggests impairment. Sensitivity and specificity of the test varies from 56 to 79 percent and 58 to 91 percent respectively (Strauss et al., 2006).

Watson: The Watson method (Watson et al., 1993) uses a "pre-drawn" clock. In this method, the examinee is instructed to draw the numbers on the clock face but is not asked to draw the hands (examinees are not asked to indicate a time). The clock is evaluated only for the positioning of clock numbers. For scoring the clock, the clock face is divided into four quadrants. The fourth quadrant (i.e. 9 to 12 o'clock) receives the greatest weight. If any error is committed in quadrant 4, a score of 4 is given; quadrants 1 to 3 receive a score of 1, if any error appears in them. Scores in this method range from 0 (least impaired) to a maximum of 7 (most impaired). An examinee is considered normal if he scores from 0 to 3 and a score greater than or equal to 4 is considered abnormal (i.e. suffers from dementia). Storey et al. (Storey et al., 2001 cited in Strauss et al., 2006) report a sensitivity of 69 percent and a specificity of 44 percent when the method was used on a sample of geriatric medical outpatients.

Freedman: The Freedman method (Freedman et al., 1994) uses a "free-drawn" clock and the time specified is 6:45. The total score is 15 with higher scores indicating better performance. This method of scoring was developed by Freedman from a normative study of 348 individuals ranging in age from 20 to 90years. The clock drawings were analysed to determine the responses that were present at a high or low rate, respectively. From the youngest age group, a subset of descriptors was selected if they occurred or did not occur, in at least 90 percent of the subjects. The authors then evaluated the list of descriptors to ascertain which were considered characteristic of a "good" or "bad" clock. The definition of "good" and "bad" clock was determined through consensus. While evaluating, the following question was asked: "Would the presence or absence of this item in any way contribute significantly to a "good" or "bad" clock?" for example, the data showed that at least 90 percent of the examinees who drew arrow heads, drew them on both hands of the clock (10 percent or less drew only one arrow head, either on the minute hand or the hour hand). The authors decided by consensus that although drawing one arrow head was unusual, such a clock was not considered as a significantly "bad" clock; therefore, this item was not considered as a critical item. Ultimately, 15 critical items were determined. Thus, the scoring system consists of 15 critical items that constitute a maximum score of 15 i.e., a score of 1 for each item if it is in the affirmative (see Appendix A for detail). The 15 critical items have been divided between 4 categories: Contour (2 critical items), Numbers (6 critical items), Hands (6 critical items), Center (1 critical item).

The different scoring systems not only have differing sensitivities with regard to the detection of dementia but also differ in their ability to monitor progression of cognitive deterioration. Powlishta, et al., (Powlishta, et al., 2002) tested 15 cognitively normal subjects (CDR=0), 25 very mild dementia subjects (CDR=0.5), 21 mild dementia subjects (CDR=1) and 14 moderate (CDR=2) to severe dementia(CDR=3) subjects. Six clock drawing methods (Manos, Mendez, Sunderland, Rouleau, Pfizer, AD Cooperative Society) were used on all individuals. They found out that the very mild dementia group did not differ significantly from the normal group, with regard to their mean clock drawing abilities regardless of the clock scoring method used to evaluate their performance. Also, they failed to find a significant difference between the mild and the moderate to severe groups. They concluded that for the detection of very mild dementia, the clock drawing test did not appear to be a good screening instrument. But, when Heinik et al. (Heinik et al., 2002a) used the Shulman and the Freedman methods on subjects with mild (CDR=1) DAT (Dementia of the Alzheimer's type) and moderate (CDR=2) DAT, they reported that performance of the moderate DAT subjects on the Freedman method was significantly worse than mild DAT subjects, while as performance on the Shulman method was similar between the two groups.

According to Esteban-Santillan et al. (Watson et al., 1993 cited in Esteban-Santillan et al., 1998), if an examinee is unable to accurately denote time by use of hands of the clock, it is a significant indication of cognitive impairment and should be considered as a more serious error than "primary artistic, visual, or spatial difficulties". They further added that their results suggest that very high positive and negative predictive values for early Alzheimer's disease are provided by hand placement items, on the clock drawing test; therefore, when the Clock Drawing Test is used in dementia screening evaluations, the accurate placement and depiction of hands should be given particular attention.

In general, the Tuokko, Mendez, and Shulman scoring procedures are quite sensitive. However, they have a modest specificity; specificity of the Mendez method is particularly poor (Strauss et al., 2006). Besides, in the Mendez method, full credit can be obtained without the hands indicating correct time which does not count towards a merit of the method as a highly diagnostic feature of the clock drawing strategy is not evaluated appropriately. Brodaty and Moore (1997) compared three scoring methods (Shulman, Sunderland, and Wolf-Klein) in a memory disorders clinic and reported that the Shulman's method performed best and even it was superior to the MMSE. In a sample of geriatric outpatients, Storey et al. (Storey et al., 2001 cited in Strauss et al., 2006) compared five scoring methods (Shulman, Sunderland, Wolf-Klein, Mendez, and Watson); they found that dementia was predicted more accurately by the Shulman and Mendez methods than by the Sunderland, Wolf-Klein or Watson methods. Using older adults with and without dementia, Tuokko et al. (Tuokko et al., 2000) compared a number of scoring methods and reported high intra-rater reliability for the Wolf-Klein and the Shulman scoring methods and concluded that the relatively lower inter-rater reliability coefficients of these methods most likely reflected the subjective nature of these methods; they also reported lower sensitivities and specificities for the Wolf-Klein and Watson scoring methods. Besides, in the Watson method, the examinees are instructed to draw the numbers on the clock face but are not asked to draw the hands (examinees are not asked to indicate a time). Hence there is no representation of a highly diagnostic feature of the clock drawing strategy. The Sunderland scoring system also has a lower sensitivity and specificity (Strauss et al., 2006).

Suhr et al. (Suhr et al., 1998) evaluated the quantitative and qualitative performance of stroke patients and normal controls on six clock drawing systems and reported the highest interrater reliability (0.98) for the Freedman scoring method. Heinik et al. (Heinik et al., 2004) in a study of DAT (dementia of the Alzheimer type) and VD (Vascular dementia) patients reported significant correlations of 0.8 and higher between the Freedman and the

Shulman scoring system; also, performance of the VD patients was significantly worse than the DAT patients on the Freedman method, but not on the Shulman method. Heinik et al. (Heinik et al., 2002b) made a study to see if the clock drawing test could differentiate between dementia of the Alzheimer's type and vascular dementia; they reported that only the Freedman clock drawing total score and hands subscore were statistically different between the two groups and hypothesized that it was the presumed sensitivity of the Freedman method to impaired executive functioning that it was able to differentiate between the two groups, because in vascular dementia impaired executive functioning is more pronounced compared with DAT (dementia of the Alzheimer's type). The Freedman's hands subscore is the "most abstractive-executive oriented element" in the method. In another study, the use of the clock drawing test was studied in Schizophrenia and control subjects (Herrmann, et al., 1999); they reported that the Schizophrenia patients had significantly worse performance on the Freedman method compared to controls despite performance being equal on the MMSE. Also, the Freedman method of clock drawing is a "free-drawn" method and hence constructional abilities (i.e. visuoconstructional task) are required for the reproduction of the visual aspects of a clock without model of a clock being available to copy from. Shulman (2000) admits that the Freedman method because of its length and complexity appears well suited for research purposes.

Freedman (Freedman et al., 1994) highlights that for the screening of cognitive impairment, although the clock drawing test is a sensitive screening measure, however it is not intended as a tool for the diagnosis of any specific disorder, such as Alzheimer's disease. It rather exposes the "deficits due to dysfunction in specific brain systems" that may result from a broad array of neurological disorders.

## 3.4 Rey-Osterrieth Complex Figure Test

This test was developed by Rey in 1941 and was subsequently elaborated by Osterrieth in 1944 (Strauss et al., 2006). Osterrieth in 1944 developed its 36-point scoring system for scoring the accuracy of its reproduction and was translated into English by E.M. Taylor in 1959 and is in wide use ever since (Hamby et al., 1993). The Rey-Osterrieth Complex figure is reproduced below (Figure 3.1) and comprises of 18 standard items (see Appendix-A).

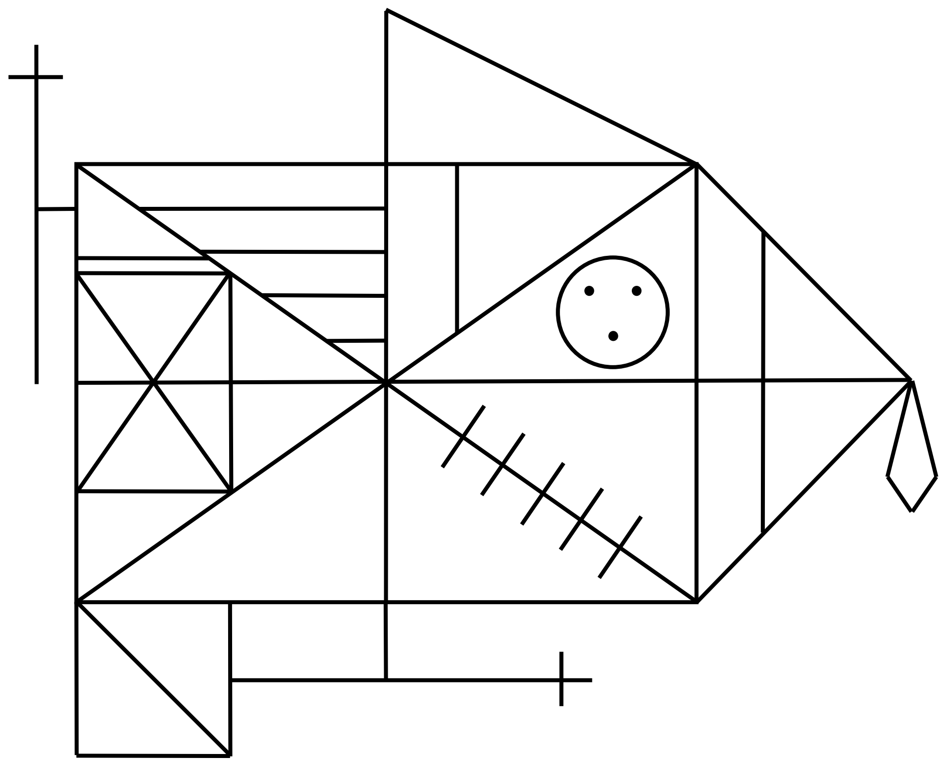


**Figure 3.1** The Rey-Osterrieth Complex Figure (Lezak et al., 2004).

In this test, the examinee is given a drawing of the Rey-Osterrieth Complex Figure and asked to make a copy of the figure. After the examinee has drawn the figure, the original and drawn figures are taken away and he is asked to reproduce the drawing from memory “immediately” and/or after a certain “delay”. Typical measures of performance include the copy score (it indicates how accurately the figure has been copied and is a measure of visual-constructional ability) and “immediate” and/or “delayed” recall scores (which is a measure of the amount of information retained over time). In most administrations, the examinees are not forewarned when they are being given the copy instructions that they will have to reproduce the figure from memory (Lezak et al., 2004). Knight et al. (Knight et al., 2003 cited in Strauss et al., 2006) made a survey of members of the International Neuropsychology Society and found out that the most frequently used interval between the copy and "immediate" recall was from 0 to 5 seconds and the average interval between the copy trial and the delayed trial was 27 minutes (SD=14 minutes) while as 30 minutes was the most frequently used interval. The test originally developed by Rey in 1941 involved the copy trial and a "delayed" recall 3 minutes later (Strauss et al., 2006). Lezak (Lezak et al., 2004) report that some examiners follow Osterrieth's convention and use a delayed recall interval of 3 minutes.

According to Strauss et al. (2006), the test assesses visual-spatial construction ability and visual memory. They further elaborate (Meyers & Meyers, 1995a; Waber & Holmes, 1986 cited in Strauss et al., 2006) that the test assesses a broad range of cognitive processes which include planning skills, organizational skills, and problem solving strategies. Perceptual, motor and episodic memory functions are also assessed. According to Chervinsky et al. (Chervinsky et al., 1992 cited in Strauss et al., 2006), the general notion is that if adequate performance is to be achieved on the test, the precise cognitive operations that must be exercised include visual perception, visual-spatial organization, motor functioning and memory (on the recall condition). Overall, correlational and factor analytic studies show that validity of the test as a measure of visual-constructional ability (copy) and memory (recall) but not executive function (Strauss et al., 2006). Impairments in visual memory also contributes to poor decision making (based on incorrect input) thereby increasing drivers' risk of errors, crashes and injuries (Rizzo & Kellison, 2004) . The performance difference between the immediate and 3-minute delayed recall scores is very small; also, overall recall performance is not affected by the length of delay chosen (15, 30, 45, or 60 minutes), provided the delay is not greater than 1 hour. Most forgetting takes place within the first few minutes after copying (Meyers & Meyers, 1995a; Berry & Carpenter, 1992 cited in Strauss et al., 2006). Although, in normal subjects virtually no forgetting takes place between the immediate recall and 20- or 30-minute delayed recalls (Tombaugh & Hubley, 1991). Also, if an examinee takes the immediate recall test, it will affect his performance on the delayed recall (Lezak et al., 2004).

Qualitative aspects of performance on the test can also be judged by noting the process or strategy an examinee uses while copying /recalling the figure (Ruffolo et al., 2001). Henceforth, some examiners record the strategy adopted by the examinee, while he is copying by either the colour pencil method or the flowchart method. This enables the examiner to capture the sequential process used while drawing. Because, the manner in which the figure is copied (i.e., strategy and organization) will have a significant bearing on recalling the figure. Both copy and recall accuracy scores tend to be correlated with qualitative copy scores (Strauss et al., 2006). According to Lezak et al. (2004), largely examinees who approach the copying task conceptually i.e. drawing the overall configuration of the design and then as a second step filling in the details, recall the figure much better than participants who copy the details one by one even if it is done systematically (i.e., going from left to right or top to bottom). This difference may be due to the fact that when items are processed in pieces, there is a need to recall many more items as compared with when they are combined into conceptually meaningful units and hence demonstrates that the fragmented copy approach (piecemeal copy) is inefficient for memory storage. During the copying process, the organizational strategy (or its lack) in particular of subjects who have a lower mental ability often strongly predicts subsequent recall (Fujii et al., 2000; Dawson & Grant, 2000 cited in Lezak et al., 2004).

In the colour pencil method, when a section of the drawing is completed, the examiner hands the examinee a different coloured pencil. The order in which the coloured pencils are handed over to the examinee is noted. Pen-switching is brought about at approximately equal points in the construction of the figure and is avoided when the examinee is in the middle of drawing one of the standard 18 items (Hamby et al., 1993). Typically, three to six different coloured pens are used. In the flowchart method, on a separate sheet of paper, the examiner reproduces the drawing of the examinee (as he is copying the figure) and notes the order of lines as they are drawn by numbering and their direction by means of arrows (Ruffolo et al., 2001). Ruffolo et al. (2001) made a comparison study of the two methods and reported that the colour pencil method performed better than the flowchart method.

Particularly, age above 70 years affects performance on the copy task, however, the changes are quite subtle. Mitrushina et al. (Mitrushina et al., 2005 cited in Strauss et al.,

2006) conducted a meta-analysis and reported that there is a decline in copy performance with age and also the variability increases with advancing age. The drawings of older adults (60+ years) usually have curved lines, rounded corners, gap, and overshoots which can be classified as minor inaccuracies and are attributed to mild deficiencies in visual-spatial functions or motor problems (Hartman & Potter, 1998 cited in Strauss et al., 2006). Mean scores on the delayed recall condition also show a decline with advancing age along with increased standard deviations due to the greater heterogeneity of the older age group. In older age groups, the main mistake committed is the omission of the standard items of the drawing rather than distortion of the items (Hartman & Potter, 1998; Mitrushina et al., 2005 cited in Strauss et al., 2006).

In memory tests, the test-retest reliability is less well defined compared with other tests. A normal decrement in recall is expected if a stimulus material (here exposure to the Rey-Osterrieth drawing) is not repeated after the delay interval; but, if the stimulus material is repeated, performance is expected to be better than the performance at the initial presentation of the stimulus. In test-retest studies, a normal assumption is that that the characteristic which is being measured remains stable over time; however, both the conditions highlighted previously are in conflict with this assumption (Groth-Marnat, 2000). Ranges for some scores (e.g., copy) are restricted as most normal subjects' performance is at maximum- or near-maximum level (i.e., ceiling effect is reached) and as a result, the magnitude of the test-retest correlation coefficients is artificially reduced. Keeping these considerations in mind, Meyers & Meyers (Meyers & Meyers, 1995b cited in Strauss et al., 2006) evaluated test-retest reliability of normal subjects for scores having sufficient range after a retest interval of 6 months and obtained a coefficient of 0.76 for immediate recall and 0.89 for delayed recall.

Strauss et al. (2006) report adequate to high interrater and intrarater reliability (coefficients greater than 0.80) for total scores obtained with scoring according to the criteria of Osterrieth (1944) and E.M. Taylor (1959) (or variants of these criteria). In another instance, they report a high interrater reliability for total scores (greater than 0.90). Lezak et al.(2004) also report an interrater reliability coefficient of greater than 0.90.

No significant relationship was found between Rey-Osterrieth Complex Figure Test and language measures (Strauss et al., 2006). For individuals with a history of central nervous system health problems that are known to affect memory and executive system, the test has shown sensitivity. Such health problems include Alzheimer's disease, Parkinson's disease, Huntington's disease, Korsakoff's disease, cocaine and poly drug abuse, seizure disorders, head injury, medial temporal lobe damage, Ischemic vascular dementia and anterior communicating artery aneurysm (Strauss et al., 2006). Lezak et al.(2004) also report that the test bears sensitivity to mild neuropsychological impairments that are found in a variety of clinical populations. We must bear in mind that the majority of individuals who are healthy are able to draw the figure without major distortions and obtain high scores, on the copy condition. Therefore, categorizing a high copy score as "superior" does not make any sense; however, a low score on the copy condition is cause for concern and has significant clinical implications (Mitrushina et al., 2005 cited in Strauss et al., 2006). Strauss et al. (2006) comment that for the interpretation of recall scores, the adequacy of the initial copy must be considered. If the initial copy is poor it might indicate disrupted encoding (e.g., due to visual-perceptual or organizational problems) and if recall is poor, it may suggest disrupted storage.

### 3.4.1 Rey-Osterrieth Complex Figure Test: Test Procedure & Scoring

The detailed testing procedure is at Appendix-A. The measures of performance include a copy score and 3-minutes delayed recall scores. However, in the copy condition the colour pencil method (as explained earlier) is used to capture the sequential process while drawing. The quantitative scoring of both the copy condition and the 3-minute delayed recall condition is done according to the 18-item, 36-point scoring system developed by Osterrieth in 1944 (Osterrieth, 1944) that was translated into English by E.M. Taylor in 1959 (Taylor, 1959) (see Appendix-A for detail). Some examiners also record completion time. Although, several scoring systems have been devised, the method developed by Osterrieth continues to be the most commonly used method (called the Rey-Osterrieth method) (Lezak et al., 2004; Groth-Marnat, 2000). The figure comprises of 18 scorable items (see Appendix-A). Each item is numbered to facilitate scoring. Each item receives between 0.5 and 2.0 points depending upon accuracy, distortion and location. If an item is correct and is placed properly it receives 2 points, if it is correct and is placed poorly it receives 1 point. If an item is distorted (or incomplete but recognizable) and is placed properly it receives 1 point, if it is distorted (or incomplete but recognizable) and is placed poorly it receives 0.5 points. If an item is absent or is not recognizable, it receives zero points. Therefore, the maximum score that can be achieved is 36. The criteria proposed by Osterrieth (Osterrieth, 1944) clearly specify the items that are to be scored but still considerable latitude exists with regard to classifying items as “distorted” or “misplaced”; therefore, a number of authors have proposed more explicit criteria for scoring the figure according to the Rey-Osterrieth method (Strauss et al., 2006). Strauss et al.(2006) report that the criteria developed by L.B. Taylor (reproduced in Spreen & Strauss, 1991) are stricter than those of E.M. Taylor (Taylor, 1959). For clarity and consistency in scoring, Groth-Marnat (2000) also recommends using the Taylor’s Guidelines that are reproduced in Spreen & Strauss (1991) for most purposes (see Appendix-A for Taylor’s criteria / guidelines).

## 3.5 Dichotic Listening Test

Studies have been conducted in the past to explore the relationship between driving performance and experimental measures that gauge the efficiency of focusing and switching attention. For example, Kahneman et al. (1973) conducted a study in which a test of auditory selective attention (the Dichotic Listening Test) was given to 117 professional bus drivers aged 22 to 32. In the dichotic listening test given to the subjects, a series of digits and letters were presented at the same time to each ear. Each test consisted of two parts. In the first part of the test, drivers were told to concentrate attention on one specific ear (i.e. left or right ear, this is the relevant ear) and to report (speak aloud so that the examiner notes them down) only the digits appearing on that specific ear (specific channel). In the second part of the test, they were told that when they hear a “particular” tone, they have to switch attention to the other channel (other ear) and report digits from that channel (on 50 percent of the second parts, the same ear is the relevant channel). In the beginning of each part (i.e., part one and part two) there is a tone; if the examinees hear a "low" tone that indicates that the left ear is the "relevant" ear in that part, if they hear a "high" tone that signifies that the right ear is the "relevant" ear in that part. In both parts, subjects have to only report digits from the relevant ear. The rate at which the letters/digits were spoken was 2 letters/digits per second. Table 3.1 is an example of the schematic of one such Dichotic Listening Test.

**Table 3.1** An example of a Dichotic Listening Test (Note: Each pair containing letters/digits is presented simultaneously to both ears).

| | Left Ear | Right Ear |
|---|---|---|
| Part One | | |
| Pair 1 | V | Z |
| Pair 2 | P | 6 |
| Pair 3 | J | U |
| Pair 4 | 1 | I |
| Pair 5 | W | 9 |
| Pair 6 | M | X |
| Pair 7 | 8 | E |
| Pair 8 | T | 2 |
| Pair 9 | 7 | H |
| Pair 10 | Q | 0 |
| Pair 11 | A | 3 |
| Pair 12 | L | N |
| Pair 13 | C | 5 |
| Pair 14 | 4 | K |
| Pair 15 | S | R |
| Pair 16 | B | F |
| Part Two | | |
| Pair 1 | Y | D |
| Pair 2 | 4 | 2 |
| Pair 3 | 8 | 0 |
| Pair 4 | 7 | 9 |

Essentially the test was used to gauge three categories of errors in the realm of selective attention:

1. Number of errors committed in reporting the digits from the relevant ear (relevant channel) (omission errors). This measure indicated the extent to which there was inability to focus attention on the relevant channel.
2. The number of digits reported from the irrelevant channel (intrusion errors). This measure indicated a person's vulnerability to distraction.
3. Since 50 percent of the second parts involved switching to the other channel, thus the number of errors in (1) and (2) as explained above following the switch in relevant channel in the second part of the test were coded as switching errors (switching errors).This measure indicated failure of the attention mechanism responsible for switching.

Correlation coefficients of 0.29 and 0.31 were obtained between the number of errors committed in reporting the digits from the relevant channel and the number of digits reported from the irrelevant channel respectively with vehicle crashes over a one year period. And a correlation coefficient of 0.37 was obtained between the number of switching errors and crashes over a one year period. Similar findings have been replicated in other studies. In an even earlier study by Gopher & Kahneman (1971), the dichotic listening test was applied to a highly pre-selected group of cadets (of high-performance aircraft) in the Israel Air Force and a significant correlation of 0.36 was found with a three-level criterion in pilot training. For predicting different criteria of proficiency in flying high-performance aircraft, the test had promising validity. Also, pilots of high-performance interceptor and attack aircraft had significantly better performance (on part two of the test) than pilots of transport and slower jet aircraft. They finally concluded that in driving / flying under normal conditions and in the second part of the dichotic listening test, the requirement to reorient attention is a common feature. The ability to switch attention as measured by the dichotic listening test is an indicator of overall performance since in the driving and flying tasks, the operators of the machines (car & plane) do not passively wait for orienting signals but have to rapidly switch attention between the stream of events taking place. In the pilot's case, this requires a high level of attention and integration of information from several sources. His task includes monitoring numerous displays inside the cockpit (relevant to ground location, vertical situation, and energy management etc.), response to ground communication and control of complex flight manoeuvres (Gopher, 1982). In the case of driving, the task is one of monitoring a cluttered driving environment (in urban areas) for different developments / events / traffic regulatory signs etc. and switching attention from low priority events to high priority events while at the same time he may have to monitor the instrument panel for data relevant to speed etc. Since the driving environment is dynamic, new events/ scenarios keep on appearing and the driver is in a continuous process of switching attention between different events / stimuli based on their changing priorities. Therefore, while driving, successive events call for the rapid switching of attention. Also, the driver has to avoid interference from distracting sources of information and has to divide resources properly. Intuitively, one would think that how is it possible that the scores on the dichotic listening test (a test dependent on auditory modality) could be correlated with crashes, when driving is apparently a predominantly perceptual task. To elaborate on this aspect, Avolio et al. (1985) conducted a study in which the auditory selective attention test (the Dichotic Listening Test) and a visual selective attention test (developed by the author on the model of the auditory selective attention test and constructed to equal as a visual counterpart of the auditory selective attention test), were given to seventy two drivers (aged from 28-59). The authors found out that all three categories of errors (i.e., omission, intrusion and switching errors) on the auditory selective attention test (Dichotic Listening Test) were significantly correlated (in the predicted direction) with individual accident rates. However, in the visual selective attention test, only omission errors and switching errors were significantly correlated (in the predicted direction) with individual accident rate. The intercorrelations among the errors of the auditory selective attention test and the visual selective attention test were positive and significant. Also, it was observed that the correlations between the switching errors of the auditory selective attention test and the visual selective attention test had the highest correlations (among the test battery) with individual accident rate. Finally the authors concluded that since both measures of selective attention (i.e., both modality-specific measures i.e. auditory as well as visual) were correlated with each other and were also correlated with an external task (i.e. the individual accident rate), therefore both may be tapping a central cognitive construct / domain, that is modality-free. Gopher (1982) gave the Dichotic Listening Test to 2000 flight cadets in the Israeli Air Force and found switching errors had the highest correlation with success in flight training and the dichotic listening test did add to the predictive value of the test battery employed for pilot selection and concluded that since there were high intercorrelations between types of errors and the dichotic listening test predicted performance in the "primarily visually loaded piloting task" , therefore the dichotic listening test in essence was tapping a single central cognitive function that was modality free. Parasuraman & Nestor (1991) give a list of other studies (along with relevant data in tabular format) that show significant correlation starting from 0.3 to 0.4 between measures of driving performance (using accident rate or closed-course driving performance index) and at least one measure of selective attention. One primary fact that surfaces from these studies is that the largest correlations were obtained from the switching error category of selective attention. In this context, Kahneman et al. (1973) at the conclusion of their study suggested that to reorient attention from an earlier state of attention to a channel/stimulus is more difficult than to initially apply focused attention from an uncommitted waiting state.

According to Lachman et al. (1979), it is in the brain that attention is accomplished and not in the senses (e.g., vision, hearing etc.) as studies using visual stimuli and auditory stimuli gives the same results; however, it is much easier to use the auditory stimuli, whereby different messages are presented to the two ears. Both messages are transmitted to the brain and selectivity can be accomplished centrally in the brain (as to which message to inhibit). Therefore, we opted to develop an auditory task (i.e. the Dichotic Listening Test) rather than a visual task keeping in view equipment costs and simplicity of testing procedure as the Dichotic Listening Test can be administered without the use of a computer.

Perry and Hodges (Perry & Hodges, 1999 cited in Duchek & Balota, 2005) suggest that prior to any deficit in language or visuospatial abilities in DAT (Dementia of the Alzheimer's Type), attention is the first non-memory aspect of cognition that exhibits decline. Elaborating further, Baddeley et al.(2001) point out that in the early stage of DAT, sustained and focused attention tend to be relatively preserved; however, across several tasks, deficits in selective attention have been reported (Duchek & Balota, 2005). Also, according to Parasuraman & Nestor (1991) accident risk is critically related to selective attention and the ability to shift selective attention. Particularly, a decline in the efficiency of visual selective attention is seen in the early stages of DAT and hence may contribute to impaired driving performance (Parasuraman & Nestor 1991). Parasuraman & Nestor (1993) also report that the ability to switch attention is also correlated with driving performance in normal individuals.

### 3.5.1 Dichotic Listening Test: Design

In the Dichotic Listening Test (Gopher & Kahneman, 1971; Kahneman et al., 1973; Gopher, 1982; Avolio et al.,1985; McKenna et al., 1986) there were 48 Dichotic tests. Each test consisted of two parts. In part one, there were 16 pairs of digits/letters presented simultaneously to the two ears. The pairs were presented at a rate of 2 pairs per second. The digits ranged from 0 to 9 and the alphabet from A to Z. In part two, there were three pairs of digits that were preceded by 0,1, or 2 additional pairs of letters. The presentation rate in part two was the same as that in part one i.e., 2 pairs per second. In the beginning of each part (i.e., part one and part two) there was a tone; when the examinees heard a "low" tone that indicated that the left ear was the "relevant" ear in that part, if they heard a "high" tone that signified that the right ear was the "relevant" ear in that part. In both parts, subjects had to only report aloud the digits from the relevant ear. A low tone was designated by sounding a tone having a frequency of 250 hertz and a high tone by a frequency of 2500 hertz. The typical layout of a test is as follows: First there is a silence of 5 seconds then there is an announcement of test no. (e.g., Test No.1) on both ears. The examinee has to repeat aloud the test no. Then there is a silence of 2.5 seconds followed by a either a high tone or a low tone of 0.5 second duration, for part one of the test. The tone is presented monaurally to the relevant ear (i.e., if it is a low tone, it is presented to the left ear and if it is a high tone, it is presented to the right ear). After a silence of 1.5 seconds, the simultaneous presentation of 16 pairs of digits/letters begins at a rate of 2 pairs per second. After part one, there is a silence of 1.5 seconds and is followed by a either a high tone or a low tone of 0.5 second duration, for part two of the test. The tone is always presented monaurally to the ear that is relevant in part one of the test. Then after a silence of 1.5 seconds, there is a simultaneous presentation of three pairs of digits that are preceded by 0,1, or 2 additional pairs of letters. There are some additional features (constraints) of the test that have been outlined below:

1. On 50 percent of the tests, the tone (cue) in part one is opposite to that of part two.
2. In part one, there is no pair that has only digits i.e. a pair has a digit and a letter such as letter-digit or digit-letter.
3. In part one, the relevant ear (channel) had either two or four digits, whereas the irrelevant channel always had six digits. In part two of each test, there were always three digits on the relevant ear and three on the irrelevant ear.
4. In part one the relevant digits are different from the irrelevant digits and also in part two, the relevant digits are different from the irrelevant digits. Furthermore, taking parts one and two together, all relevant digits in a test are different and so are also all irrelevant digits.
5. In the 16 pairs in part one, the digits are randomly placed between position 2 and 14 and at least one letter is between successive digits.

24 test were fabricated by factorially crossing 2 (instruction tones for part one i.e., either left ear relevant or right ear relevant) × 2 (instruction tones for part two i.e., either left ear relevant or right ear relevant) × 2 (number of relevant digits in part one i.e., either 2 or 4) × 3 (number of letter pairs preceding the digit pairs at the start of part two i.e., either 0, 1 or 2). Therefore $2 \times 2 \times 2 \times 3 = 24$ tests were fabricated. In fabricating the tests, the digits for the relevant and irrelevant parts were chosen randomly through a random number generation program and then the order of digits within a test was also randomized. The configuration of the numbers was chosen so as to fulfil all the aforementioned constraints. Finally, the order of the factorial design was also randomized to obtain a test sequence which included 24 tests. Instead of a 250 hertz low tone, a 150 hertz low tone was used to make the discrimination between the low tone and the high tone more obvious. Also, the presentation rate of the digits / letters was set at 1 digit/letter per second instead of 2, because the test involved older individuals who found the 2 digit/letter per second rate to be too fast to grasp. This was determined after a pilot sample of younger and older individuals was tested at various presentation rates.

The Dichotic Listening Tests performed by Gopher & Kahneman (1971), Kahneman et al. (1973), and Gopher (1982) comprised of 48 Tests. However, to reduce the duration of testing time, we used 24 tests as per the factorial design outlined earlier. Boer et al. (1997) gave 48 Dichotic Listening Test to Royal Netherlands Air Force aviator applicants and Royal Netherlands Navy air traffic control applicants and reported higher validity for the 24 test protocol as compared to the 48 test protocol and finally recommended the shorter version of the Test i.e. 24 Dichotic Tests. They also reported the adoption of the shorter protocol by the Royal Netherlands Air Force for their pilot selection program. Avolio et al. (1985) in their study of individual differences in information-processing ability as a predictor of motor vehicle accidents also used 24 Dichotic Listening Tests instead of 48 Tests.

### 3.5.2 Dichotic Listening: Recording & Equipment

In the Dichotic listening test, one of the critical issues is the synchronization of the stimulus (digits/letters) onset between the channels (ears) so that a precise temporal alignment is achieved between the two ears. According to Berlin (Berlin, 1977 cited in Davey, 1987) even if there is a difference as small as 10 milliseconds between the ears in the start of a stimulus it will affect the advantage accorded to a particular ear. Therefore to fabricate the Dichotic Listening test it was necessary to use a computer to generate tapes wherein there is exact control of temporal parameters relevant to the technique. For this purpose we used a sound editing software called AUDACITY® . This software is a digital audio editor and is used for: (1) recording live audio (2) convert tapes and records into digital recordings or CDs (3) edit MP3/WAV sound files (4) cut, copy, splice, and mix sounds together (5) change the speed or pitch of a recording and other tasks.

For recording purposes, a researcher with a clear voice, good pronunciation and whose first language was English was employed. Using AUDACITY®, we recorded: (a) “Test 1”, “Test 2” till “Test 24” (b) “0”, “1”, “2” till “9” (c) “A”, “B” till “Z”. To avoid confounding a “zero” with a “O” (i.e., pronounced as OOOO), we avoided recording a “O”. In the second step, we examined the wave forms of each of these spoken elements, listened to them and truncated them so as to get rid of the excess audio signature at the two ends and then labelled them. As a final product, these elements had to sound just about right. If there was an ambiguity in the pronunciation of an element, we requested the researcher to rerecord that element again. Finally we made sure that the time duration of each element (except “Test 1”, “Test 2” etc) did not exceed 0.5 seconds to prevent an overlap with the following element. Although this software is not user friendly for recording the Dichotic listening test (but then there is no other software to do so), we had to painstakingly assemble each of the elements along with silences, keeping in view the temporal duration of each element to the millionth of a second to get a presentation rate of 1 digits/letters per second and to achieve synchronization of stimulus (digit/letter) onset between channels so that the temporal alignment between the ears was to a millionth of a second. The individual tests were then appended to each other in series using the software. Test 1 to Test 12 were combined to form one series, while as Test 13 to Test 24 were combined to form another series. We also recorded instructions and sound samples of “low tone” and “high tone” so that the examinees become familiar with the tones. All these digital files were in a format that is specifically used / recognized by AUDACITY® ; therefore, to make these files compatible with Microsoft Windows® Media Player, these files were converted into WAV format using the software. The Dichotic Listening test prepared for this research is placed on the accompanying CD in WAV format. A good quality stereo headphone set is also required to take the test.

### 3.5.3 Dichotic Listening Test: Scoring

Score on the test is the total number of errors. Essentially the errors can be categorized into three categories:

- Omission errors: Number of errors committed in reporting the digits from the relevant ear (relevant channel). This measure indicates the extent to which there was inability to focus attention on the relevant channel.
- Intrusion errors: The number of digits reported from the irrelevant channel. This measure indicates a person’s vulnerability to distraction.
- Switching errors: Since 50 percent of the second parts of the tests (i.e., 12 tests out of 24) involved switching to the other channel (the other ear), thus the number

of Omission and Intrusion errors (as explained above) following the switch in relevant channel in the second part of those 12 tests are coded as switching errors. This measure indicates failure of the attention mechanism responsible for switching.

To facilitate scoring of the Dichotic Listening Test, a score sheet was prepared (see Appendix-A). In this score sheet all the relevant digits for part one and part two of each test were written in bold font in one line. The irrelevant digits were noted in a line above the line of relevant digits. Tests wherein the switching of the relevant channel (from left to right or right to left) took place were given a shade to distinguish them from the rest, as switching errors are only noted in these tests (these tests were twelve in number). The examinee-dictated digits were noted on this sheet of paper in front of each test. To facilitate ambiguity-free scoring, we bifurcated the examinee-dictated digits into two groups (i.e., part one and part two of the test) by examining the actual relevant and irrelevant digits. For Omission errors, we noted the number of relevant digits in part one that were omitted by the examinees in each test and then added them up for all the tests; these were the Omission errors. We divided this by 72 and multiplied it by 100 to get percentage Omitted errors. We divided by 72 because in the 24 tests, 12 tests had 2 relevant digits and 12 had 4 relevant digits in part one.

For Intrusion errors, we inspected the examinee-dictated digits/letters in part one and compared them with the irrelevant digits/letters in part one. The number of common digits/letters between these comparisons was the intrusions. These were added for all 24 tests to obtain Intrusion errors. We divided this by 144 and multiplied it by 100 to get percentage Intrusion errors. We divided by 144 because in the 24 tests, each test had six irrelevant digits in part one, so the maximum number of intrusions that one could make was six in each test. For the switching errors category, since 50 percent of the second parts of the tests (i.e., 12 tests out of 24) involved switching to the other channel (the other ear), thus the number of Omission and Intrusion errors (as explained above) following the switch in relevant channel in the second part of these 12 tests were coded as switching errors. For these errors we had to consider only the 12 tests where the switching in the channel took place (i.e., Test number 2,3,5,10,13,14,17,19,20,22,23 and 24). The switching errors for these tests were added to get switching errors. We divided these by 72 and multiplied it by 100 to get percentage Switching errors. We divided by 72 because in the 12 tests under consideration, in each test there were 3 relevant digits in part two that could be missed and there were also three irrelevant digits that could have intruded.

To get the total number of errors, the number of Omission, Intrusion and Switching errors were added for all 24 tests. Percentage Total errors were calculated by dividing this number by 288 (i.e., 72+144+72 = 288).

## 3.6 Paper Folding Test

In the Paper Folding Test, subjects are mentally required to fold a piece of paper, imagine that a hole is punched through the folded paper and then visualize as to what the paper would look like when it is unfolded. In the test, successive displays of figures show a square piece of paper that undergoes from one to four overlapping folds sequentially. Then a figure shows one or two holes punched through (such that it passes through the whole thickness of the paper) the folded paper at a particular location. There are five figures on the right hand side; one of these figures shows the correct position of the holes when the paper is completely unfolded. The examinee has to select the correct figure from the five alternatives (i.e., A,B,C,D,E). The figure below shows a very simple example of a Paper Folding Test (single fold) (Figure 3.2). Figure 3.3 illustrates the rationale for selecting C as the correct answer, as the paper is unfolded.

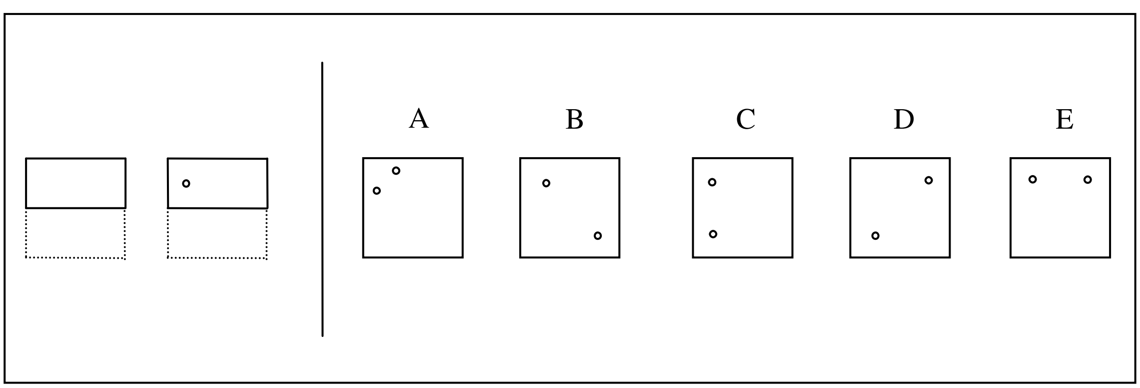



**Figure 3.2** A simple example of Paper Folding Test (Ekstrom et al., 1976)

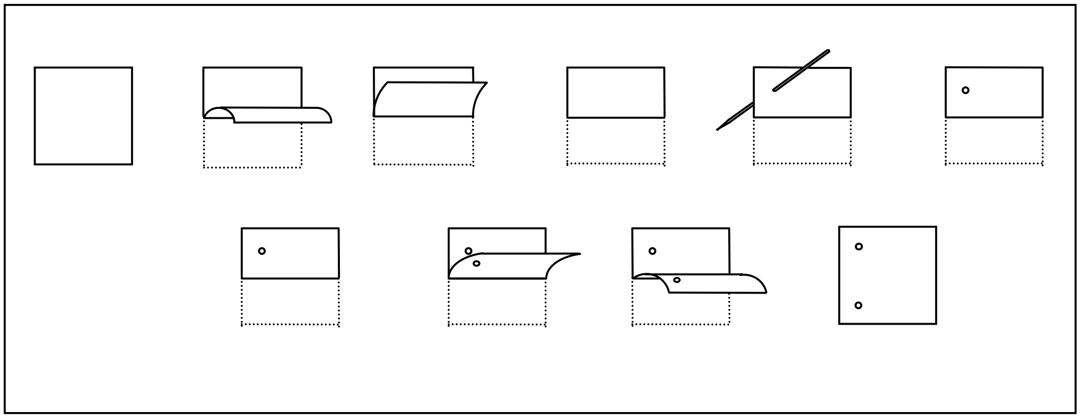


**Figure 3.3** Showing unfolding of paper after punching of hole (Ekstrom et al., 1976).

According to Ekstrom et al. (1976), the Paper Folding Test measures spatial visualization ability. Spatial visualization refers to the mental manipulation of spatial information to determine how a given spatial configuration would appear if portions of that configuration were to be rotated, folded, repositioned, or otherwise transformed (Salthouse et al., 1990). De Raedt & Ponjaert-Kristoffersen (2000) classify the test as a visuo-spatial test with a working memory component because visuo-spatial information has to be stored in working memory while at the same time another mental operation is being performed. The test involves a mental reversal process that relies on working memory (De Raedt & Ponjaert-Kristofferson, 2001). The simultaneous storage and processing of information is a critical aspect of working memory. For a task that assesses working memory, a necessary and important feature is the maintenance of some information while that information or some other information is being processed upon (Baddeley, 1986 cited in Salthouse et al., 1989). Impairment of working memory is associated with ageing according to many reports (Craik & Rabinowitz, 1984 cited in Salthouse et al., 1989). The paper folding test also puts considerable demand on the central executive, because in order to solve the problem, a sequence of internal spatial transformations have to be exercised on the figure; therefore, there is a requirement that the examinee manages task specific goals and subgoals and also different cognitive processes have to be scheduled and coordinated (Hegarty et al., 2000).

Ekstrom et al. (Ekstrom et al., 1976 cited in Blajenkova et al., 2006) report a test-retest reliability of 0.84 for the paper folding test and Salthouse et al. (1990) report a reliability of greater than 0.82. In another study, Salthouse et al.(1989) report a reliability of 0.86.The test also shows important age effects (Craik & Salthouse, 1992). Craik & Salthouse (1992) tested 383 adults between the ages of 18 and 80 years. The scores on the test were converted into z scores (standardized scores) based on the scores of the young adults (18-29 years old). The distributions of the scores were plotted for each age cohort in 10 year increments (i.e., 20s, 30s, 40s, 50s, 60s, 70s). The distribution plots showed a clear tendency for the entire distribution of scores in each cohort to shift to lower levels as the age increased. A correlation coefficient of 0.42 was obtained between the paper folding test and age.

De Raedt & Ponjaert-Kristofferson (2000,2001) made a study of 84 car drivers between the ages of 65 and 96 years who were referred for a fitness-to-drive evaluation. The relationship between their self-reported accidents and seven neuropsychological tests was investigated. Their driving performance was also evaluated on a road test. Among the neuropsychological tests, the Paper Folding Test was the best predictor of accidents at cross roads. The authors explained that in order to make predictions about the time-to-arrival of upcoming vehicles, information on initial position of vehicles must be retained in working memory. Therefore, visuo-spatial function with working memory plays an important role in left-turn (right turn in Britain) performance, primarily because the judgement of speed and distance of oncoming vehicles is involved. The situation is further exacerbated because multiple oncoming and crossing vehicles require that drivers coordinate their turn movements into gaps that are multiply constrained because of the directional split of traffic and that some times more than one lane is in each direction

Also, according to Guerrier et al. (1999), the importance of working memory in driving cannot be overemphasized because, for example, while making a left turn (Right turn in Britain), based on the dynamic changing situation of oncoming traffic and driver's own vehicle capabilities (e.g. acceleration), information has to be processed /operated on, stored, retrieved and decision made (Guerrier et al., 1999).

### 3.6.1 Paper Folding Test: Test Procedure & Scoring

Detailed Testing procedure and problem items are at Appendix-A. Ponjaert-Kristoffersen (2000) used the paper folding test with only one fold and the examinees had to choose between only two alternatives. However, we designed the test ourselves and included a total of 16 test items with 4 items each of one, two, three and four overlapping folds. Also, we gave the examinees 5 alternatives (i.e., A,B,C,D,E) to choose from on each test item. As the number of folds increase, choosing the correct answer becomes more difficult. Salthouse et al.(1989) found that as the number of folds increased mean accuracy decreased for all trials. They concluded that this decrease in accuracy with increase in folds showed an inability on part of the examinees to preserve relevant information when other information was being processed and thus appeared to represent a failure of the working memory. Based on these finding it was intuitive that the 16 test items should not be scored equally as they varied in the number of folds. Therefore, we devised a scoring system wherein the score of each test item (if correct) was proportional to the number of folds in that item. Therefore, the one fold item was allocated 2.5 marks (if correct), the two folds 5 marks, the 3 folds 7.5 marks and the 4 folds 10 marks. Since there are 4 items each of the one, two, three and four folds, if an examinee answers all of them correctly he gets 100 marks (i.e., $2.5 \times 4 + 5 \times 4 + 7.5 \times 4 + 10 \times 4 = 100$). To discourage guessing, some marks are subtracted from the total score, for incorrect answers; this is in accordance with the recommendation of Ekstrom et al. (1976). Therefore, guessing is not to the advantage of candidates unless they eliminate one or more of the answer alternatives as wrong. Since each item has 5 alternatives, the probability that a guess will be correct is one fifth; therefore if an item is incorrect, one fifth of the marks of that item will be subtracted from the total score. For example, if an examinee answers all items correctly except one 3 fold item and one four fold item his net score will be 79 (i.e., correct marks – incorrect marks = $(2.5 \times 4 + 5 \times 4 + 7.5 \times 3 + 10 \times 3) - (7.5 \times 1/5 + 10 \times 1/5) = 82.5 - 3.5 = 79$). This protocol adopted for determining the number of marks to be subtracted for incorrect answers is quite in conformance with the procedure adopted by Blajenkova et al. (2006) and Hegarty et al. (2000).

According to Hegarty et al. (Carroll, 1993 cited in Hegarty et al., 2000), in the Paper Folding Test the difficulty of the items to be solved primarily constrains an individual's performance and hence the time limit is relatively liberal. In accordance with these recommendations, we set a time limit of 10 minutes for the 16 items of the test.

## 3.7 Useful Field of View (UFOV) Test

Sanders (Sanders, 1970 cited in Edwards et al., 2005) was the first researcher to put forward the concept of useful field of view. The visual field area over which information can be acquired in a brief glance without eye or head movement was termed "functional visual field" by the author. This concept was introduced by researchers because they found out that although standard acuity and perimetric tests relevant to the visual field were able to diagnose eye diseases, they failed to predict/explain the difficulties encountered by older individuals in performing everyday tasks that required the use of peripheral vision. The term Useful Field of View was first used by Ball et al. (1990) and is ultimately now associated with a specific test that is administered by a personal computer and is developed by Visual Awareness, Inc., Birmingham, AL, USA.

Scientists and clinicians from a wide spectrum of disciplines agree that information processing speed slows with age and is due to the fact that neural transmission speed declines and the nervous system of older individuals shows signs of slow recovery from the effects of stimulation. Worsening of the cognitive function with advancing age is primarily attributed to this slowing which also manifests itself prominently in the performance of older people on complex activities (Visual Awareness Inc., 2002). Research in the areas of visual information processing and cognitive ageing conducted during the 1980s and 1990s refined the concept of UFOV and its measurement. Research showed (Scialfa et al., 1987; Bergen & Julesz, 1983 cited in Edwards et al., 2005) that to search an area having a particular size for a target, either a serial search strategy or a parallel search strategy is adopted which is related to the duration of the stimulus (the exposure of the area) as well as the target/distractor similarity or dissimilarity (corresponds to conspicuity of the target) i.e., given a constant duration, the more conspicuous the target the greater are the eccentricities at which it can be detected compared with a less conspicuous target and given a constant conspicuity, the longer the stimulus duration, the further are the eccentricities at which the target can be detected compared with shorter durations of the stimulus. By varying the stimulus duration and the conspicuity of the peripheral target and the identification of a central object, an individual's useful field of view can be manipulated. These guiding principles were used in the development of the UFOV which included 3 subtests (subtest 1, 2 and 3) that measured an individual's processing speed across increasingly complex visual displays.

The original version (Standard version) was administered via the Visual Attention Analyzer which had a 20 inch touch screen and included a chin rest that was situated 23.5 cm from the screen (Clay et al., 2005). This test required 20 to 30 minutes to administer. In the Standard version, in the divided (subtest 2) and selective attention (subtest 3) subtests, the peripheral targets appeared randomly at eccentricities of 10, 20 and 30 degrees along eight radial spokes (Clay et al., 2005). The duration of displays varied between 40 and 240 milliseconds (Owsley et al., 1998a). For each subtest, the scores in the Standard version are expressed in terms of the percentage reduction of a maximum 30-degree radius field (Owsley et al., 1998a) and range between 0 and 30 percent reduction. Since the factors known to impact the size of the useful filed of view are independently assessed by each of the subtests, thus the three subtest scores are added to obtain an overall measure of reduction in UFOV (Edwards et al., 2005). Thus in the Standard version, this measure ranges between 0 and 90. Higher scores indicate poorer performance. A score of 0 percent reduction on a subset means that the individual was able to correctly radially localize targets at all eccentricities at the minimum stimulus (i.e. 40 milliseconds) duration and a score of 30 percent reduction on a subset means that the individual cannot reliably localize the peripheral targets presented at the 10 degrees eccentricity and the maximum stimulus duration (240 milliseconds). Ball et al. (1993) in a study showed that a UFOV (relative to a 30 degree radius) reduction of greater than 40 percent provided the best single cutpoint for separating high risk and low risk drivers.

To expand UFOV's application, two new and briefer versions of the UFOV were developed that are administered via a personal desktop computer. Beginning in 1998 the test is now administered using the PC versions (Clay et al., 2005). One version uses a touch screen monitor and the other uses a mouse to register responses. Scores on the two

PC versions correlate very highly with each other (Edwards et al., 2005). The test conducted under both versions takes about 15 minutes. Peripheral target eccentricity and display duration are the most critical factors affecting the size of UFOV. In the original test (Standard version), the peripheral target eccentricity was manipulated for a fixed display duration. The size of UFOV was measured at that duration and then the process was repeated at a faster duration. In the PC versions, the eccentricity of the peripheral target is fixed and the display duration is manipulated. In the PC (personal computer) versions, the peripheral target appears at approximately 11 cm from the center of the fixation box (center of the screen) and the duration of displays varies between 16.67 (say 17) and 500 milliseconds (Edwards et al., 2005). The PC versions do not have a chin rest but examiners are instructed to seat examinees at about 18 to 24 inches from the screen (Visual Awareness Inc., 2002). The monitor size used is 17 inches (Visual Awareness Inc., 2002) for the PC versions. The examinees may be given a short guided-practice program for using the mouse. Also, prior computer experience is not required to use either version (i.e., Standard version or PC versions) and also prior computer experience does not enhance performance in any systematic way (Edwards et al., 2005). In the PC versions, scores for each subtest represent the duration of displays in milliseconds at which examinees could correctly perform each subtest 75 percent of the time. The software automatically adjusts the length of stimulus presentation in milliseconds as needed for each of the three subtests. The stimulus presentation time for an item is shortened if the previous two responses are correct and is lengthened if the previous response is incorrect. This process of tracking the perceptual threshold is continued until a stable estimate of 75 percent correct is calculated (Visual Awareness Inc., 2002). The length of time that the stimulus is displayed will get shorter and shorter (provided correct responses are made). The software will measure the point at which the examinee fails to accurately see the information displayed on the screen and therefore there will be a particular time when the test becomes impossible for everyone (Visual Awareness Inc., 2002).

In the PC versions, the scores represent the display duration in milliseconds at which the examinee correctly performs each subtest 75 percent of the time. A metric for inter conversion has been prepared (see Edwards et al., 2005) for converting percent reduction in UFOV score (used in the Standard version) to subtest threshold in milliseconds (used in the PC versions) and vice versa. Visual Awareness Inc. (2002) provides tables for detailed interpretation of either scoring system. The three subtests measure visual processing speed, ability to divide attention and selective attention abilities (Myers et al., 2000). However, performance on the three constituent subtests is non-independent as the speed of visual processing is relevant to all three subtests and in subtest 2 and 3, attention abilities are relevant (Owsley et al., 1998a). The UFOV test is a quantitative index of visual processing speed and divided and spatial attention (St.Louis et al., 2005).

In the UFOV test, in the first subtest the examinee is to identify a target (silhouette of a car or truck) that is presented in a central fixation box and is presented for varying lengths of time. In the second subtest the examinee has to identify a central target (silhouette of a car or truck presented in the central fixation box) and also to radially localize a simultaneously presented target (silhouette of a car) displayed in any one of eight radial positions on the periphery of the computer monitor. The third subtest is identical to the second subtest except that the peripheral target (Silhouette of a car) is embedded in distractors (47 triangles of the same size and luminance as the peripheral target). When unsure of the correct answer, examinees are encouraged to guess as the test would not go forward unless a response is made.

UFOV is not a measure of visual field sensitivity and UFOV results should not be translated to percentage visual field loss (traditionally assessed by standard vision perimetry tests). A reduction in the useful field of view may be due to visual field loss or due to impairments in attention or both of the factors (Visual Awareness Inc., 2002). Visual sensory and cognitive deficits can occur in older people together or separately. In a study (Ball et al., 1990 cited in Ball et al., 1993), it has been shown that UFOV shrinkage can occur even in older subjects with excellent visual field sensitivity. Most driving license issuing agencies have put a lot of emphasis on the assessment of visual acuity. However, as reported by Ball et al. (2006), although, several large scale sample studies have attempted but have failed to show a link between visual deficits (including several indices of visual function) and crash involvement, thus showing that visual function alone is a poor predictor of driving performance. Conventional measures of visual field assess visual-sensory sensitivity, whereas UFOV is linked to higher order processing skills, such as rapid visual-processing speed, selective and divided attention. Owsley et al. (1995) also demonstrated that performance on the UFOV relies on higher-order cognitive abilities as well as visual sensory function. Sekuler et al. (2000) describes the deterioration in the useful field of view as a decrease in efficiency of extracting information from a cluttered scene. They also found out that the deterioration of UFOV starts early in life (by 20 years or younger).

In a study by Ball et al. (1993), to identify visual factors associated with increased vehicle crashes in elderly drivers, useful field of view test had high sensitivity (89 percent) and specificity (81 percent) in predicting the crash history of elderly drivers. It was observed that older drivers with significant deficits in useful field of view were six times more likely to be involved in accidents during the past 5 years. Significant correlation was obtained between crashes and eye health status, visual sensory function and chronological age, but these parameters were poor at distinguishing crash-involved drivers from crash-free drivers. In fact, According to the author, these tests (sensory tests, such as visual acuity and peripheral field sensitivity) do not reflect the visual complexity of the driving task, and are more relevant to clinical diagnoses and assessment of ocular disease/vision loss. The driving environment is quite complex, where vehicle control has to be negotiated in a cluttered environment through the simultaneous use of both central and peripheral vision to process both primary (high priority) and secondary (low priority) visual tasks and where the prediction of important events in time and space is unpredictable. Therefore, simple visual sensory tests fail to capture the visual demands of driving. In the Ball et al. (1993) study, the subjects whose visual acuity was better than 20/20, 43 percent had a useful field of view (UFOV) reduction of greater than 40 percent (the threshold amount) and 41 percent of the subjects who had a useful field of view (UFOV) reduction of greater than 40 percent (the threshold amount), showed an average loss of visual field sensitivity of less than 2.5 dB.

The UFOV is not a test that assesses reaction time. If performance on subset 1 is poor, it may signify poor central vision, processing speed, attention, working memory, or a combination of these factors (Visual Awareness Inc., 2002). Generally if performance on subset 1 is poor, performance on subset 2 will also be poor as all the features of subtest 1 are present in subset 2 and subset 2 has an additional target in the periphery. Subtest 1 is sensitive in identifying individuals that are severely impaired (Edwards et al., 2005).

Edwards et al. (2005) in their study found out that the UFOV test was significantly correlated with age. Visual Awareness Inc. (Visual Awareness Inc., 2002) also cites three relatively early studies that indicated that in general the useful field of view appears to shrink with age. In a population based study, approximately one third of the older subjects had a 40 percent or greater reduction in useful field of view (Rubin et al., 1997 cited in Owsley, et al., 1998a). However, there is considerable variability and many older subjects' performance is at par with that of young college students. Therefore, not all older individuals are affected in the same manner. Reger et al. (Reger et al., 2004 cited in Clay et al., 2005) found subtest 2 to have the highest correlation with at-fault crashes. Subjects having diseases such as Alzheimer's and Parkinson's and traumatic-brain injured subjects have shown abnormal performance on the UFOV test (Fisk et al., 2002; Rizzo et al., 2000; Uc et al., 2003 cited in St.Louis et al., 2005). Duchek et al. (1998) carried out a study using three groups: (a) healthy control group (58 subjects) (b) very mild DAT (dementia of the Alzheimer's type) group (49 subjects) (c) mild DAT group (29 subjects). All participants were administered the 45 minute Washington University Road Test (WURT). A correlation coefficient of - 0.56 was obtained between the UFOV scores and the driving scores such that poorer driving performance was significantly related to greater reduction in UFOV. Also, UFOV scores increased (became worse) as the severity of dementia increased. The average reduction in UFOV of healthy controls, very mild DAT subjects and mild DAT subjects was 29, 34 and 75 percent respectively. These averages were based on administering the UFOV test to 28 control subjects, 21 very mild DAT subjects and only 6 mild DAT subjects. The authors concluded that mild DAT subjects appeared to exhibit a deficit in their ability to "process peripheral target information amidst visual distractors, while simultaneously monitoring a central task".

Owsley, et al.(1998a) carried out a prospective cohort study of 294 drivers aged 55 to 87 years with three years of follow up from 1990-1993 to identify whether measures of visual processing ability, including the Useful Field of View test, are related with crash involvement in older drivers. In their study, visual attention and visual processing speed was assessed using the Useful Field of View test. They found out that older drivers with a reduction of useful field of view of 40 percent or greater were more than twice as likely to have experienced an accident during the three years of follow-up. In another study of older drivers performed at the University of Alabama at Birmingham, Owsley et al. (1991) found that deficits in information processing ability as measured by the useful field of view test and deficits in cognitive abilities were related to crash involvement as recorded by the state. By incorporating their parameters in a model they were able to explain 20 percent of variance in crash involvement (as recorded by the state). Further, it was reckoned that older drivers with poor scores on the UFOV or exhibiting poor cognitive status had 3-4 times more accidents (of any type) and 15 times more intersection crashes than subjects without those problems. Goode et al. (1998) conducted a study of 239 older drivers ranging from age 55 to 85 and above and found that the UFOV test was most strongly related to crash involvement (state recorded at-fault accidents over the previous five years) among all cognitive tests. At a cut-off value of 40 percent reduction in UFOV score, they obtained a sensitivity and specificity of 86.3 and 84.3 percent respectively. Studies made by other authors have also found the UFOV test to be predictive of accidents in older drivers (Ball & Owsley, 1993; Rizzo et al., 1997; Sims et al, 1998). Owsley et al. (1998b) made an exploratory study (from 1985 to 1990 in the state of Alabama) of 78 drivers involved in injurious crashes, 101 drivers involved in non-injurious crashes and 115 drivers not involved in any crash during the same period. The drivers' age ranged from 55 to 87 years. They reported that elderly drivers having reduction of UFOV greater than 40 percent were at least 20 times more likely to be involved in an injurious crash compared with subjects with no or more minor reductions in UFOV and concluded that visual processing impairments and eye conditions were more prevalent in older drivers involved in injurious vehicle crashes. The UFOV task depends on an individual's speed of processing, divided attention and selective attention performance and taps abilities that are vital to driving at the attentive (serial) and pre-attentive (parallel) levels. Stimulus and task features that are critical for driving are indeed incorporated in the useful field of view test. A driver with reduced UFOV may perform as if he or she has tunnel vision and yet he or she may not show any abnormality on standard vision perimetry tests, which place more emphasis on maximal estimates of sensory function vis-à-vis attention effects (Rizzo, 2004). The UFOV is quite relevant to situations where a person has to be aware of peripheral objects e.g., such as at intersections. Intersections can be particularly demanding for the driver as they require a series of complex visual perceptions and decision making processes. Visual attention at the preattentive level plays a very crucial role as it is used to quickly direct one's attention to highly salient visual events occurring in a cluttered scene (such as busy traffic intersections) e.g., when a vehicle approaches in peripheral vision (from the cross road).

Visual Awareness Inc.(2002) reports a test-retest reliability coefficient of 0.88 after testing 70 participants aged 65 years and older. Edwards et al. (2005) in their study evaluated the test re-test reliability coefficients for the UFOV PC mouse version and the UFOV PC touch screen version as 0.884 and 0.735 respectively. Also, a high correlation coefficient (0.916) was obtained between the scores of both the PC versions. Correlation coefficients of 0.658 and 0.746 were obtained between the Standard version and the mouse and between Standard version and the touch screen version respectively. The coefficients 0.916 is slightly larger than the coefficients 0.658 and 0.746 because both the PC versions use the same scoring scale, the same size monitor and the possibility of score values has a larger range. The authors finally concluded that practical use of both the PC versions (mouse and touch screen) in clinical practice / evaluation can be adopted due to the sufficient magnitude of their reliability and validity coefficients.

### 3.7.1 UFOV Test: Scoring and Testing Procedure

Presently, only the PC versions of the test are being marketed. We used the PC version that employs the mouse to make responses. For detailed testing procedure, the UFOV manual may be consulted (Visual Awareness Inc., 2002). However, before the test begins, examinees are informed that the test would not go forward unless a response is made, therefore if they are unsure of the correct answer, they should guess. Also, during the course of the test, there will be a point when the test becomes impossible and therefore, they should not be alarmed if they cannot recognize the displayed presentation at that time. Test scores for each subtset are generated by the computer. For interpretation of the scores, the UFOV manual may be consulted (Visual Awareness Inc., 2002). Also, Appendix-A contains a table for inter conversion (adopted from Edwards et al., 2005) of percent reduction in UFOV score (used in the Standard version) to subtest threshold in milliseconds (used in the PC versions) and vice versa.

## 3.8 Summary

This chapter has described the cognitive attributes that are relevant / crucial for the driving task and has then come up with the selection and detailed description of a battery of cognitive tests that are appropriate in gauging decline in relevant cognitive constructs. Multiple cognitive domains are called upon when an individual is negotiating driving scenarios/situations and many of these cognitive constructs themselves are interrelated and interact with each other. The coordination of several ongoing processes including attention, perception, memory (declarative, procedural, and working), and executive functions (decision making and implementation) is required for safe driving. A meta-analysis of 27 primary studies showed that visuospatial skills and attention were most helpful in screening at risk drivers. In case of head injury or generalized brain damage, tests that measure visual memory, executive abilities, spatial awareness and attention were deemed significant predictors of driving skills. In the early stages of DAT (Dementia of the Alzheimer's type), subjects show impairments in the ability to switch or disengage attention, while as they retain the ability to focus attention to a spatial location. The ability to switch attention is also correlated with driving performance in normal individuals.

During the course of normal ageing, some cognitive functions e.g. spatial orientation and perceptual speed have been found to decline. Also, in normal cognitive ageing, there is a general decrement in the speed with which information is processed, the efficiency with which new information is acquired also decreases, there is cognitive inflexibility and the working memory shows signs of reduction. The ability to control spatial attention lies on a continuum from healthy young adults to the young-old, through the old-old to individuals with mild DAT.

Six cognitive tests (Trail Making-B Test, Clock-Drawing Test, Rey-Osterrieth Test, Dichotic Listening Test, Paper Folding Test and the UFOV Test) were selected. This chapter also describes these tests in detail (including the cognitive domains tapped by each test) and their testing and scoring procedures. These tests were selected as they covered key cognitive domains (as highlighted above) necessary for safe driving, were diverse (e.g. paper-and pencil tests, computerized and listening tests), were reliable and were quite sensitive to the effects of ageing and to a range of diseases that are well known to impair driving performance. Tests that measured vocabulary and general information were not included. Each cognitive test did not reflect a pure measure of a single cognitive domain, but rather each test tapped more than one cognitive domain. Also, each test only partially tapped a specific domain. There was more than one test that tapped the same domain that was very critical relevant to the driving task e.g., there was more than one test that tapped visuospatial ability and attention (because these were highly crucial domains relevant to driving).

# 4 Driving Simulation

## 4.1 Introduction

The objectives of this chapter were to describe the simulation equipment, design of simulation drives, refinement/fine tuning of drives, sample of drivers, the experimental procedure and the various driving performance parameters and their processing. Usually, the criteria to assess driving ability are based on driving measured in on-road conditions. However, researchers (Christie, 1996; Vagverket, 2001 cited in Patomella & Kottorp (2005)) have questioned the ability of on-road tests to highlight unsafe driving behavior. Usually, the normal on-road test is insufficiently challenging to identify risky driving behavior due to cognitive impairment (Vagverket, 2001 cited in Patomella & Kottorp (2005)). Also, due to the non-standardized nature of normal road tests, each driver is not subjected to the same opportunities to commit errors and so the scoring of driving maneuvers may be of doubtful validity (Fox et al., 1998). According to Fox et al. (1998), unless on-road tests are developed on the basis of solid psychometric principles, they cannot be expected to be valid barometers of driving ability, be it to assess brain impaired drivers or for ordinary driver license testing. Ulfarsson et al. (2006) in a study of factors affecting common vehicle-to-vehicle collision types using the accident database compiled by NHTSA concluded that in-vogue road tests are not practical to identify impaired drivers. Therefore, it was decided to address the research objectives through the application of Driving Simulation, and the process is described in the remaining sections of this chapter.

## 4.2 Apparatus and Equipment

A simulation drive was programmed using the STISIM® software developed by Systems Technology Incorporated, 13766 South Hawthorne Blvd. Hawthorne, CA 90250-7083.

The version of the software was Build 2.08.01 copyright © 1985-2009. Figure 4.1 and 4.2 show views of the STISIM® driving simulator.

**Figure 4.1** The STISIM driving simulator. The systems monitor to control the simulation is on the left.

The steering wheel (force feedback steering) and pedal assembly had a USB interface and were manufactured by Logitech® designated as *G25 Racing Wheel*. The Central Processing Unit was a DELL® computer with dual display video cards with 2 GB of virtual memory with an INTEL® processor operating at 2.8 giga hertz. The operating system was Microsoft windows XP Professional. There were two monitors: (1) A 16 inch Samsung SyncMaster 753DFX and (2) A 32 inch RELISYS Liquid Crystal Display monitor with a refresher rate of 60 hertz. The 32 inch monitor provided a resolution of 1024 × 768 and was the main display monitor through which drivers viewed the simulated road environment. The 16 inch monitor was used as a systems monitor to control the simulation.

**Figure 4.2** The STISIM driving simulator with the pedal assembly and the steering wheel in view.

The video card for the main monitor was *Sync Master 753DF(T)/783DF(T) Magic Sync Master AQ17DF on NVIDIA* and for the 16 inch monitor *Plug and Play Monitor on NVIDIA Geforce FX5200*. A Speaker system was provided to impart a realistic impression of sound effects such as road noise, noises emanating from the engine and instructions to the driver. The simulator was essentially a fixed base simulator with vehicle feedback provided through three modalities; these were visual feedback, audio and steering force feedback.

All the hardware was configured as per instructions in the STISIM ® Manual. The pedals were configured in order to emulate an automatic transmission i.e. the clutch pedal was made un-functional. The steering force feedback resolution was adjusted using a number of parameters. The following levels of steering adjustment parameters were adopted after a number of test drivers found them to be appropriate enough to provide a realistic sensation of steering feel and reactive force : (1) Overall Effects Strength = 25% (2)

Spring Effect Strength = 50% (3) Damper Effect Strength=15 % (4) Centering Spring Strength=100 %. The ten parameters relevant to “vehicle dynamics” controlled how the vehicle would brake, accelerate and steer; these were adjusted after input from a couple of test drivers, and were: (1) Yaw Rate Scale Factor = 65000 rad/sec/deg (2) Oversteer coefficient = 0 (3) Acceleration Limit = 0.4 g (4) Deceleration Limit = -0.8 g (5) Coefficient of Drag = 0.0001 (6) Yaw Instability = 0 (7) Speed Instability = 0 (8) Steering Deadband = 0 degrees (9) Yaw Instability Tag = 0 sec (10) Idle Throttle Setting = 0.05.

## 4.3 Design of Drives

### 4.3.1 Overview

In order to discriminate between the driving skills of cognitively impaired drivers and normal drivers it is important that drivers be presented with driving scenarios that employ driving skills at the “controlled processing” (effortful processing) level rather than those which only require “automatic processing” (Lee et al., 2002a). Automatic processes (e.g., gear changing, lane keeping, steering etc.) are acquired through many years of practice and do not deteriorate with age (Lee et al., 2002a). Also, automatic processes are fast, involuntary and place limited demands on attention capacity, whereas “effortful” or “controlled” processes are slow and capacity-demanding and are used in coping with unpredictable or unfamiliar stimulus demands. In old age, automatic routines remain relatively well preserved but older people find it very difficult to inhibit automatic processes in suddenly changing (and unexpected) situations (Rogers and Fisk, 1991). As pointed out by De Raedt & Ponjaert-Kristoffersen (2001), such situations may be encountered, for example, in rear-end accidents when the leading vehicle suddenly stops; the ability to switch from automatic to controlled processes plays a critical role in such situations. Controlled processing comes into play when routine reactions do not suffice and the complexity of the situation necessitates the use of the attention controller, the central executive (Lundqvist, 2001).

In other words, drivers should be put through complex situations / hazardous situations where specific/particular information processing stages are stressed and errors of one type or another are likely to be committed. These complex / hazardous situations should be diverse and sufficient in number so as to enable an adequate sampling of driving behaviour. Routine diverse driving maneuvers (encountered in day-to-day driving) should form the bulk with the complex/hazardous events interspersed within them. Also, driving maneuvers that older drivers find problematic or face difficulty performing need to be included. Overall, the drive should be relatively difficult and long enough to provide variability and detailed information on specific type of errors committed by drivers.

### 4.3.2 Design of Main Drive (Drive-I)

Drive-I consisted of a 21 mile drive (duration approximately 40 minutes). In order to maintain safety, negotiating the elements in the drive required situation awareness, hazard perception and decision making skills; some of these elements were:

- Controlled Hazards (pedestrian/dog intrusions, intersection intrusions, sudden-braking, intrusions with limited sight distance etc.).
- Right turns across oncoming traffic.
- Left turns (involving cyclist).
- Dangerous overtaking by opposing vehicle.
- Lane changing manoeuvres (in traffic) and lane drops.
- Stop controlled intersections (gap selection).
- Overtaking manoeuvres in the wake of a stream of oncoming vehicles.
- Transiting construction zones.
- Signalized intersections.
- Signalized intersections with dilemma zones.
- Gas station manoeuvres.
- Tracking task (boxes fallen from truck on the road, mountain S-curves).

Situation awareness in cognitive context is the ability to judge and decide, when one has to choose between alternative actions in complex/critical situations involving traffic control devices (signals, signs and markings), interacting vehicles/pedestrians and combinations of road geometry (Allen et al., 2004a). The drive was programmed using the programming language "Scenario Definition Language" (SDL) pertinent to STISIM®. The breakdown of the lane configuration was as follows:

From start to 34100 feet = 2 lane road

From 34100 to 95000 feet = 4 lane road

From 95000 to 99000 feet = 6 lane road

From 99000 to 103600 feet = residential road

From 103600 to 110400feet = 2 lane mountain road

From 34600 feet to 99000 feet and from 103900 to 110400 feet, no centreline crossing was allowed. The drive was predominantly rural with an urban flavour in certain reaches. Liberal use of horizontal and vertical curves was instituted to avoid monotony in the driving process and also to limit sight distances at specific locations so as to elicit/assess certain driver behaviour. The radii of horizontal curves were generally generous; only one or two had relatively tight radii (i.e., curves intentionally made sharper) which made them marginally drivable at the posted speed limit. Posted speed limits of 25, 30, 45, 55 and 65 mph were used in the drive. Due consideration was given to the location of critical events and measurement of performance measures in the roadway environment to ensure that such locations were conducive to the event and that the roadway geometry (i.e. for example curves, intersections and hills etc) was or was not a factor. In STISIM®, when designing dynamic events, a trigger is needed to commence a particular action from the event. A common trigger used is the headway time between the driver and the object. However, there is an option to select from a variety of triggers (e.g., distance, absolute distance, lateral distance, range, signal change etc.), and therefore quite diverse triggers were used in different events. The headway time is calculated using the driver's speed and trajectory at the time when the event is called. In setting up hazards, due consideration was given to controlling perceptual and timing variables. Some iterations were necessary to develop critical timing in situations requiring limited sight distance. To complicate lane changing, vehicles in adjacent lanes approached from behind (could be spotted in rear view mirror); lane blockages were then programmed to force the driver to manoeuvre. Safe execution of the scenario required that the driver monitor traffic, pedestrians and traffic control devices through the rear view mirrors and the windscreen and make appropriate manoeuvres at the right time. Critical events were not packed tightly together but were rather separated by quite some distance to ensure that one scenario did not spill into the other or an interaction did not take place between two adjacent events (e.g., dynamic vehicles from an upstream event carrying over into a downstream event). A significant number of benign events were introduced into the drive, but which were not a source of stimulus (e.g., a pedestrian standing by the road edge who did not attempt to cross the road as the driver approached). In the drive design, event sequence patterns were avoided so that subjects did not learn from repeated exposure.

Infrastructure consisting of static objects such as buildings, parked cars, trees, and road signs etc. was added and was thoughtfully designed giving due regard to as to how they would affect a driver's visual search pattern, perception, attention and/or driving behaviour in context of the critical event. Along the drive, several cars were also parked on the road-side shoulders with the intention that the driver should become accustomed to their presence and when these vehicles become part of a critical/hazardous event in a subsequent scenario, the driver should not be predisposed to ascribe any emergency meaning to their placement. Telephone poles at a spacing of 200 feet were placed along the road alignment to give drivers cues as to their speed.

The drive was programmed using "Scenario Definition Language" (SDL) pertinent to STISIM® . The final programming code along with PDEs (Previously Defined Events) under the folder PDE1s is attached at Appendix-C. PDEs are programs called by the main code, which also transfers some parameters to them. The Main Drive (Drive-I) file is by the name of old.evt. The overall coding of this drive took 18 months of full time work.

### 4.3.3 Design of DA (Divided Attention) and Car-Following Drive (Drive-II)

The DA and the Car-Following drive collectively designated as Drive-II, was a 14 mile drive with a total duration of 16 minutes, with about 8 minutes consumed by the DA portion. The first portion of the drive comprised of a DA (Divided Attention) task while the second portion was a Car-Following task. Both portions were in tandem with each other and there was no break in-between. In order to assess continuous measures of driving ability (e.g., speed, steering control/ lane keeping), the driver or the vehicle can be stimulated in a controlled manner e.g., by the application of wind gust, lead vehicle with a controlled velocity profile or a divided attention task etc. , and the driver's response to the stimulus measured (Allen et al., 2005). Due to the workload from competing sources, it is not possible for the driver to respond in an optimum manner to the primary task (driving) and secondary task (e.g., DA task) and one or both are bound to suffer. This trade-off can be measured and may show up as an increased reaction time, deterioration in lane positioning or speed adherence.

The DA drive was based on a two-lane roadway. There were no horizontal curves, turns, pedestrians, traffic or traffic control devices. However, fourteen vertical curves were inserted (but not at locations where the driver was supposed to immediately brake in response to a STOP sign appearing on the screen) with the intent that the vertical curves would have an effect on the speed of the vehicle and extra effort would be needed in order to maintain a constant speed while at the same time attending to the DA task. The driver had to maintain a constant speed of 55 mph. After commencement, default diamond signs appeared in the upper left and upper right corners of the screen. In a pseudo-random fashion, the diamond signs changed into one of the following signs: (1) Left triangle (2) Right triangle (3) Left down triangle (4) Left up triangle (5) Right down triangle (6) Right up triangle (7) Left horn (8) Right horn. The signs changed 58 times during the DA portion of the drive. In response to the left triangle and the down triangle, the driver was asked to press the "left divided attention" button and the "right divided attention" button was supposed to be pressed in response to the right triangle and the up triangle. For the horn symbol, the "horn divided attention" button had to be pressed. Using the configuration's input/output tab dialog box (in the software), specific buttons on the steering console were assigned to "left divided attention", "right divided attention" and "horn divided attention" . The symbols were displayed until the driver responded or the maximum time elapsed (which was 5 sec). At three pseudo-random locations in the DA portion of the drive, STOP sign appeared in the middle of the screen. The driver was asked to respond with immediate emergency braking to a stop. The brakes had to be pressed whilst the sign remained on the screen (i.e., 10 sec). The purpose of the test was to get a measure of the brake perception-reaction time.

The Car-Following portion of the drive was also based on a two-lane roadway. There were no horizontal curves, vertical curves, turns, pedestrians, or traffic control devices. A lead car appeared that the driver had to follow. The speed of the lead car fluctuated based on a pre-specified sinusoidal profile. The driver had to match the lead vehicle's speed in order to maintain a constant following distance between himself/herself and the lead vehicle. If the distance by which the driver fell behind the lead vehicle exceeded 200 feet, a warning message was pronounced approximately every 5 seconds instructing the driver that he/she was falling behind the lead vehicle and must catch up and match the lead vehicle's speed. Changes in the speed of the lead vehicle were in sinusoidal wave form having an amplitude of 10 ft/sec. The length of each cycle was 15 sec (i.e. period of the wave). The initial speed (also the mean speed) of the lead vehicle was 80 ft/sec (aprox. 55 mph). The lead vehicle speed climbed to 90 ft/sec in 3.75 sec then returned to 80 ft/sec in another 3.75 sec then plummeted to 70 ft/sec in another 3.75 sec and then climbed back to 80 ft/sec in another 3.75 sec; this pattern was repeated for the whole drive. There was no speed limit in the Car-Following portion of the drive. The final programming code for Drive-II is attached at Appendix-C by the name of DA1.evt. The overall coding of this drive took one month of full time work.

### 4.3.4 Design of Practice Drives

Two practice drives were designed, one for each drive. The practice drives were of about 12 minutes duration each. While designing the practice drive for Drive-I and Drive-II, the same principles were employed as those used in the design of Drive-I and Drive-II respectively. A flavour of all major categories of the scenarios of the actual drives was included in the practice drives. However, the trigger point parameters were altered so as to avoid a practice effect. Also, the spatial arrangement and visual appearance of scenes was altered, so that drivers were not predisposed to associate the onset of a critical/hazardous event with a particular pattern of road infrastructure and/or event. The intent of the practice drives was to develop driver's understanding of the logic used in negotiating the scenarios, to gain expertise in the use of pedals/steering etc., and to sense the acceleration/braking potential and other operational capabilities of their vehicle in the simulated environment. The final programming code for practice drive for Drive-I and Drive-II is at Appendix-C under the file names old(PRAC).Evt and DA(Prac).evt respectively. The overall coding of the practice drives took one month of full time work.

### 4.3.5 Refinement/Fine-Tuning of Drives

Testing and final refinement of the drives was essential to ensure that the drives would achieve their objectives. Two experienced middle aged drivers were given a chance to drive both drives and based on their input a number of alterations were incorporated. Some major changes are mentioned below:

- Three consecutive speed limit sign boards at a spacing of 200 feet were installed for emphasis.
- Due to an absence of cues to judge speed, the speed potential of the simulator was programmed at a maximum value of 65 mph and frequent verbal reminders were issued to drivers throughout the drive.
- Vehicles were positioned on the shoulder, in the opposing lane or adjacent same-direction lane so as to prevent drivers from avoiding certain hazardous scenarios.
- When making right turns in the face of opposing traffic, gap sizes in the opposing stream of traffic were increased.
- The sight distance to certain hazards was increased.
- Examples of all ambiguous scenarios/events where deviant behaviour was bound to occur were included in the practice drive.
- More practice of steering use was included in the practice drive.
- Inter-vehicle distance to facilitate safe overtaking was increased.
- Recorded instructions for right turn were advanced some distance upstream to prevent conflict with vehicles in the blind spot in adjacent lanes.
- The entry point to a petrol station was made more prominent through pavement colouring.

- Lane closure was incorporated to force drivers to travel in a specific lane which was relevant to the trigger point/hazard encounter.
- The importance of the number of low-speed warnings was emphasised in the instructions.
- A number of trigger point parameters relevant to scenarios/hazards were incrementally altered so as to make them more discriminatory, but not too difficult.

Refinement / fine-tuning of the code of Drive-I and Drive-II took about two months of full time work, as a number of trigger points/parameters relevant to events had to be manipulated / altered incrementally and then tested a number of times by the same two drivers. The whole process of coding, testing, refinement and re-*testing* of all drives (practice as well as actual drives) took about 22 months of full time work.

## 4.4 The Experimental Procedure

### 4.4.1 Overview and the Sample

Since drivers crash more frequently in a simulator than in real life (Fiorentino & Parseghian, 1997), it was decided to use a comparative approach to evaluate driving performance (or decline in driving competence) by testing experienced younger drivers and experienced older drivers. This would enable the identification/quantification of driving errors that are unique to decline in driving competences. For the younger age group, the statistically safest age group of 26 to 40 years was selected (Hakamies-Blomqvist, 1993) whereas for the older group the age limit was fixed as above 60 years. To minimize cross-sectional overlap in functionality, subjects between the ages of 41 to 59 years were excluded. All drivers were to have a valid UK driving license and be current drivers with at least 5 years of driving experience.

The selection of a random sample was neither feasible nor desired. Therefore, a convenient sample of volunteers (from both age groups) was sought. The selection bias introduced as a result of the non-random sample was not of much concern as it was not intended to generalize the estimated percentage of drivers (with impaired driving ability etc.) to the entire driving population i.e., we were not interested in estimating the proportion of drivers (with decrements in driving ability) in the population but rather were interested in the range of performance capabilities of drivers. Also, there were good prospects that the non-random sample may enable the testing of some “information rich” cases of older drivers which were really crucial and valuable. There have been some studies that have utilized smaller samples than the sample size in this research (Rizzo et al., 1997; Rizzo et al., 2001; Christie et al., 2001a; Radford et al., 2004 ; Nouri et al., 1987; Lew et al., 2005; Cox et al., 1998).

In total, fifty six drivers were successfully tested from both groups. Six drivers from the older group experienced simulation sickness syndrome and their testing had to be aborted. The demographic detail of the successfully-tested subjects may be seen in Table 4.1.

**Table 4.1** Demographic detail of the successfully-tested younger and older driver groups.

| Group | No. of Males | No. of Females | No. of Total Subjects | Minimum Age (yrs) | Maximum Age (yrs) | Mean Age (yrs) | Median Age (yrs) | Standard Deviation of Age (yrs) |
|---|---|---|---|---|---|---|---|---|
| Young | 14 | 14 | 28 | 26.3 | 40 | 32.3 | 32.3 | 4.4 |
| Old | 16 | 12 | 28 | 60.3 | 88.4 | 68.7 | 66.2 | 7.4 |

Approval for the research was obtained from the School Research Ethics Committee. Advertisement leaflets were distributed in the Southampton area and in a number of bowling clubs. All subjects were tested in the morning so as to avoid systematic effects of fatigue. As the whole testing protocol lasted about 4 hours, three very old age drivers (age 79.4, 82.9 and 88.4 yrs) were tested in two sessions as to avoid systematic effects of fatigue. The participants were first given a short (3 to 4 minutes) run on the practice drive (the beginning portion had S-curves, which can expose drivers prone to simulation sickness) to ensure that the driver was not prone to simulation sickness syndrome. Nausea, disorientation and ocular problems such as eyestrain, blurred vision and eye fatigue have been reported as some of the indicators of simulation sickness in fixed-base simulators (Mourant & Thattacherry, 2000). If a participant experienced the syndrome, the practice drive was immediately terminated and the participant was deemed unfit to take the simulation drive. If a participant did not feel any discomfort in the practice drive, then the rest of the protocol followed. The sample of drivers was tested over a period of 3 months.

Firstly, participants filled out a brief questionnaire (except part IV which related to post-simulation issues), then they were given the following neuropsychological tests in random order: (1) Ufov Test (2) Dichotic Test (3) Trail-Making Test (4) Rey-Osterrieth Test (5) Paper Folding Test (6) Clock Drawing Test. This was followed by a practice drive for the Main Drive (Drive-I) and then the Main Drive. This procedure was repeated for the DA and Car-Following Drive (Drive-II). Finally part IV of the questionnaire was filled out. Frequent breaks for refreshments were provided (but not in the middle of a simulation run or a cognitive test).

### 4.4.2 Instructions

Instructions for Drive-I and Drive-II are at Appendix-C. It may be highlighted that the drivers were specifically instructed to drive at the posted speed limit. If their speed fell 5 mph or more below the speed limit, a “ding” sound would remind them (every 3 seconds) and if their speed exceeded the posted speed limit by more than about 6 ½ mph, they would be issued a speeding ticket by the software (registered as a speed exceedance); both would count negatively towards driver evaluation. The number of “ding” sounds would be recorded as the number of low-speed warnings.

Since the driving speed can be used to regulate the difficulty of the driving task (Hakamies-Blomqvist, 1993), it was deemed necessary to impose the low-speed constraint; otherwise, a driver would technically complete the simulation drive at a very low speed and not experience even a single accident. Since the timing of the triggers of critical events/hazards was closely related to the speed of the driver when the event was initiated, slow speed tended to present the hazard at a lower level of intensity to the driver. Staplin (Staplin et al., 1999) has classified overcautiousness (e.g., driving slowly) as a discriminating error. Discriminating errors are potentially dangerous errors that signify degradation in driving skill. Verbal instructions were also given to avoid hazards either by braking vigorously, swerving or using a combination of both.

### 4.4.3 Data Collection

A number of driver performance measures that were generated by the simulation were collected at a frequency of 20 hertz i.e. every 0.05 seconds. The raw data were collected in ASCII code and were later converted into Microsoft Excel® format for further processing in MATLAB® . A number of driving performance parameters/statistics for relevant segments of each drive were then calculated using different algorithms. Since a very huge amount of data was generated by the simulator, it took four months of full time work to inspect and extract relevant parameters from the data files.

## 4.5 Performance Measures

### 4.5.1 Overview

Before the data were analysed, each driver's simulation file was played back and each and every off-road accident, collision, pedestrian hit, traffic light ticket, stop sign missed, centreline crossing, road edge excursion, and illegal turn was thoroughly examined to ensure that the driving error occurred as a result of driving performance decrement and not as a result of programming artefact; e.g., the vehicular traffic was "intelligent" but when it was programmed to undertake a certain manoeuvre, during the exact manoeuvre stage it lost its "intelligent" status but reverted back to being an "intelligent" vehicle once the manoeuvre was complete. An "intelligent" vehicle stopped at traffic lights, stop signs and avoided colliding with vehicles in front of them (without breaking the laws of physics). Also, for example, in some instances a pedestrian hit was registered even though the driver's vehicle bonnet had passed the pedestrian (but once the pedestrian started walking he could not stop).

### 4.5.2 Data Collection and Processing

#### 4.5.2.1 No. of Total Hazards

This parameter was the sum of No. of Off-road Accidents, No. of Collisions, No. of Pedestrian Hits, No. of Traffic Light Tickets, No. of Stop Signs Missed, No. of Illegal Turns and No. of Stops in Middle of Traffic. All these parameters are discrete events that represent a substantial risk of crashes and traffic conflicts and have been included in this category as per recommendations of a number of authors (Staplin et al., 1999; Stern et al., 2004; Allen et al., 2004b; Freund et al., 2005b; Freund et al., 2005a; Hunt et al., 1997a; Dobbs et al., 1998; Cox et al., 1998) and they signal declining driving skill. With regard to raw driving performance parameters, two of the most important parameters i.e. *No. of Total Hazards* and *No. of Low-Speed warnings* have been graphed against age and are shown in Figure 4.3. A lowess curve has been superimposed on the graphs. It may be highlighted that the Lowess curve follows the general trend in the data. Both graphs in general show an upwards trend with increasing age.

#### 4.5.2.2 No. of Low-Speed Warnings

These were the number of “ding” sounds played every three seconds when the driver’s speed was more than 5 mph below the posted speed limit.

#### 4.5.2.3 Over Speed Limit (Percent of Time)

This showed the percentage of the total driving time that the driver was more than 6 ½ mph above the posted speed limit.

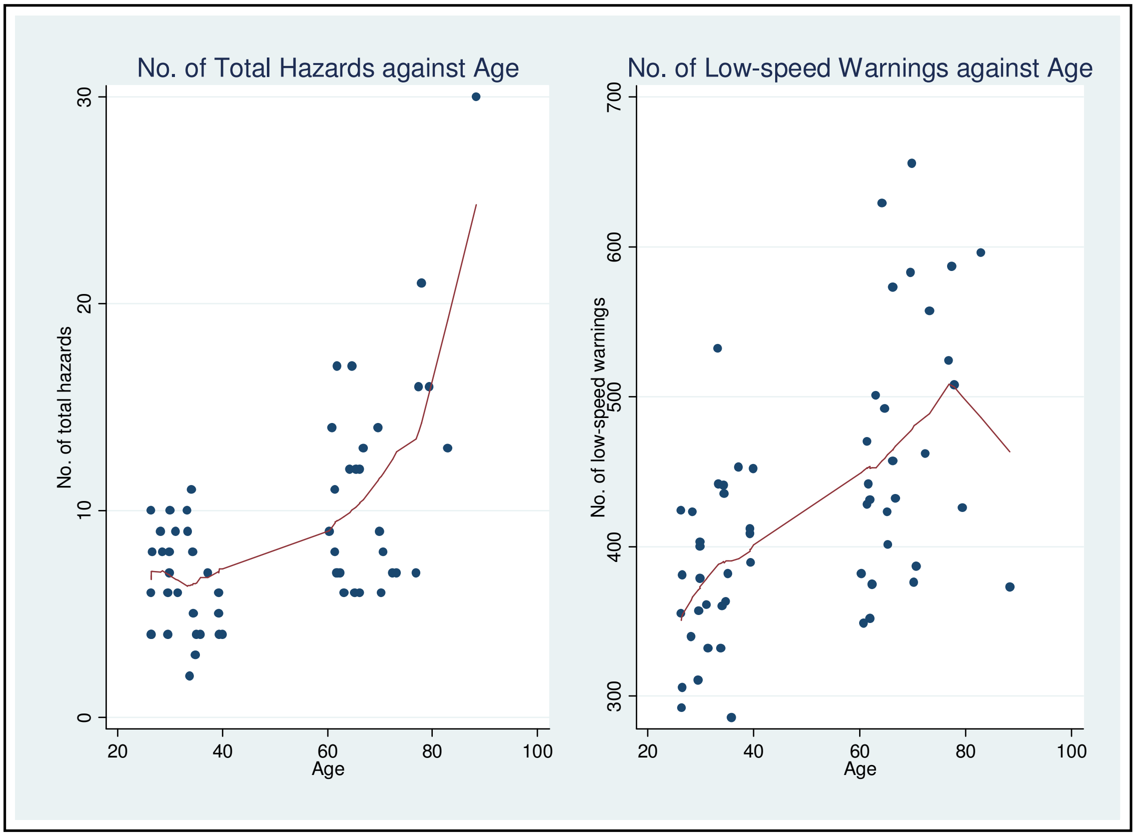



**Figure 4.3** Graphs of *No. of Total Hazards* and *No. of Low-Speed warnings* against age of drivers with superimposed Lowess curves.

### 4.5.2.4 Out of lane (Percent of Time)

This showed the percentage of driving time that a vehicle wheel was outside the left road edge or the centreline of the road. Segments of the roadway where the centreline was crossed for overtaking purposes or while evading an accident were not counted. Also, a left roadway edge crossing while veering into a petrol station was not counted.

### 4.5.2.5 Steering Reversal Rate (Mountain Drive)

It represented the number of times per minute the steering wheel was reversed by a magnitude larger than 1.5 degrees. Heading errors build up when visual attention is diverted, which are corrected by means of large and disruptive steering wheel movements. This parameter was calculated from 103900 feet to 110300 feet i.e. the mountainous portion of Drive-I.

#### 4.5.2.6 Time-To-Line crossing (Mountain Drive)

Assuming constant speed and fixed steering angle, Time-To-Line Crossing (TLC) is defined as the time required to reach the lane marking (Johansson et al., 2004); it is a time-based safety margin adopted by the driver. TLC values less than 1 second and greater than 20 seconds were ignored and the median of the remaining TLC values was computed (Johansson et al., 2004).

#### 4.5.2.7 Absolute Difference in Modulus

Modulus highlighted how closely the driver was able to match the lead vehicle's speed in the Car-Following regime. A value of 1 was ideal, whereas a value less than 1 indicated that the following driver tended to drive slower than the lead vehicle. A value greater than 1 showed that the following driver was driving faster and had to slow down on catching up. The absolute difference between 1 and the actual Modulus value was calculated.

#### 4.5.2.8 Delay (Phase Shift)

This parameter was a measure of the time it took for the following vehicle to respond to changes in the lead vehicle's speed in the Car-Following regime. A value of zero was perfect. Larger values signified that it took a longer time for the following driver to recognize and react to the lead vehicle's speed changes.

#### 4.5.2.9 Coherence

Coherence indicated how well the driver's data followed the lead vehicle's data in the Car-Following regime. It is a measure of the squared correlation between the speed signals of both vehicles (Brookhuis & Waard, 1994). A value of 1 showed perfect coherence.

#### 4.5.2.10 No. of Correct DA Responses

This indicated the number of times the driver was able to press the correct divided attention button in response to one of the eight possible signs out of 58 instances.

#### 4.5.2.11 No. of DAs with No Response

This indicated the number of times the driver did not respond at all within 5 sec to the sign change that occurred out of a total of 58.

#### 4.5.2.12 Reaction Time DA Task

This indicated the average reaction time in seconds of the driver measured from the time when the sign first appeared to when the driver pressed the divided attention button.

#### 4.5.2.13 Standard Deviation of Reaction Time DA Task

This indicated the standard deviation of the reaction time of the driver in seconds measured from the time when the sign first appeared to when the driver pressed the divided attention button.

#### 4.5.2.14 Reaction Time to Stop Sign

This indicated the average (of three readings) of the reaction time in seconds and was measured from the time the STOP sign first appeared on the screen to the time the driver's foot hit the brake pedal.

#### 4.5.2.15 Absolute Difference in Speed DA Task

This indicated the absolute difference in mph between 55 mph and the average speed taken in four segments in the DA portion of the drive where there was no interference from the appearance of the STOP signs on the screen.

#### 4.5.2.16 Standard Deviation in Speed DA Task

This indicated the standard deviation of speed in mph taken in four segments in the DA portion of the drive where there was no interference from the appearance of the STOP signs on the screen.

#### 4.5.2.17 Absolute Difference Lane Position DA Task

This indicated the average of the absolute difference in feet between the center of the vehicle and the centre of the lane in the Divided Attention portion of the drive.

#### 4.5.2.18 Standard Deviation Lane Position DA Task

This indicated the standard deviation in feet of the absolute difference between the center of the vehicle and the centre of the lane in the Divided Attention portion of the drive.

#### 4.5.2.19 Absolute Difference Lane Position Car-Following Task

This indicated the average of the absolute difference in feet between the center of the vehicle and the centre of the lane in the Car-Following task.

#### 4.5.2.20 Standard Deviation Lane Position Car-Following Task

This indicated the standard deviation in feet of the absolute difference between the center of the vehicle and the centre of the lane in the Car-Following portion of the drive.

#### 4.5.2.21 Steering Reversal Rate DA Task

It represented the number of times per minute the steering wheel was reversed by a magnitude larger than 1.5 degrees in the Divided Attention portion of the drive.

#### 4.5.2.22 Steering Reversal Rate Car-Following

It represented the number of times per minute the steering wheel was reversed by a magnitude larger than 1.5 degrees in the Car-Following portion of the drive.

#### 4.5.2.23 Time-To-Line Crossing DA Task

It indicated the Time-To-Line Crossing (TLC) value in seconds for the DA portion of the drive. TLC values less than less than 1 second and greater than 20 seconds were ignored and the median of the remaining TLC values was computed (Johansson et al., 2004).

#### 4.5.2.24 Time-To-Line Crossing Car-Following

It indicated the Time-To-Line Crossing (TLC) value in seconds for the Car-Following portion of the drive. TLC values less than less than 1 second and greater than 20 seconds were ignored and the median of the remaining TLC values was computed (Johansson et al., 2004).

## 4.6 Questionnaire

A simple questionnaire was devised to gauge driver behaviour in particular circumstances, self assessment of their driving ability. It consisted of five parts: (I) Driving Skills (II) Driving Behaviour (III) About Computing (IV) Simulation (V) Personal Information. Part-I gauged driver's self assessment of their various driving skills. Part-II focused on their driving behaviour in various circumstances. Part-III was about their commuting practices and accidents/penalty points. Part-IV focused on their assessment of the driving simulation runs. Part-V was for soliciting personal information i.e. age, sex, contact, etc. As the research developed, it became clear that whilst the questionnaire could provide insights into the attitudes of drivers, it was not relevant to the main thrust of the research and was not therefore considered further. However, a copy of the questionnaire is at Appendix-B (on a CD).

## 4.7 Summary

This chapter has described the driving simulation process involved in this research. Justification has been given as to why it was decided to meet the research objectives through driving simulation rather than on-road tests. The Simulation apparatus has been described in detail along with certain values of vital parameters that were adopted. The psychometric principles employed in the design of the simulation drives have been outlined and the design elements of both drives (Drive-I Drive-II) along with the practice drives have been described in detail. Drive-I is the main drive while as Drive-II is the Divided Attention and Car Following Drive. The drive testing and refinement process has been described.

The criterion used in the selection of sample has been ascertained. Two groups of drivers were tested in this research. Group-I was the experienced younger driver group (age 26-40 years), whereas group-II was the experienced older driver group (age > 60 years) with at least 5 years driving experience and a current valid UK driving license. The experimental procedure consisting of the six major neuropsychological tests (i.e., Ufov, Dichotic, Trail-Making, Rey-Osterrieth, Paper Folding, Clock Drawing) and the driving simulation protocol has been described.

Data collection consisted of collecting 24 driving performance parameters at every 0.05 seconds interval. The 24 driving performance parameters relevant to Drive-I and Drive-II have been described in detail. The Computation of driving performance indices from these 24 parameters has been described in the next chapter.

# 5 Driving Performance Indices

## 5.1 Introduction/Overview

The objectives of this chapter were to develop three different driving performance indices using the concepts of Cronbach's Alpha Reliability Coefficient using Item Analysis and then to derive their corresponding weighted versions through the use of Principal Components Analysis. Driving performance was objectively quantified through driving performance indices that were based on the driving performance parameters collected in Drive-I and Drive-II of the simulated drive cycles. Since there was no external criterion (e.g., driving instructor's assessment) against which to judge driving performance, therefore the concept of "Scale Development" had to be employed. The underlying phenomenon or construct that we were interested in was "driving performance". Since the construct is not directly observable, a collection of items (i.e. driving performance parameters) were combined into a composite score, because the items (i.e. values of the driving performance parameters) were *caused* by the construct (i.e. driving performance) (DeVellis, 2003). However, the scale should be homogeneous i.e., the items should tap different aspects of the same construct (driving performance) and the homogeneity also ensures that the scale is internally consistent (a set of items are internally consistent if they share common variance and tap the same underlying construct (Spector, 1992)). It would not be logical to add items to form a total /composite score if they measured different attributes. The homogeneity condition has two implications (Streiner & Norman, 2003): (1) There should be moderate correlation between the items (2) Each item should correlate with total score formed from the composite of the items. If one item is correlated highly with another, then the second item would add little additional information to the construct and be redundant. This can result in a possible loss of content validity as different/diverse aspects of the construct would fail to be represented and the scale would be narrow in scope. The methodology for the development of different driving performance indices and their rationale is described in the remaining sections of this chapter.

## 5.2 Cronbach's Alpha

A scale should measure a construct in a reproducible fashion. Reliability in a general sense is a measure of the reproducibility of a set of scores under differing conditions and shows the consistency or stability of a measuring instrument (Lyman, 1998). There are different forms of reliability (Allen &Yen, 1979). The most general method of finding estimates of reliability through internal consistency is the Cronbach's Alpha (Kaplan & Saccuzzo, 2001). Cronbach's Alpha is a measure of internal-consistency reliability that provides a lower-bound estimate of a test's reliability based on a single administration (Allen &Yen, 1979; Streiner & Norman, 2008). The formula for Cronbach's Alpha (Bland & Altman, 1997) is:

$$\alpha = \frac{\boldsymbol{k}}{\boldsymbol{k}-\mathbf{1}}\left(\mathbf{1}-\frac{\sum \boldsymbol{S_i}^{\mathbf{2}}}{\boldsymbol{S_T}^{\mathbf{2}}}\right) \tag{5.1}$$

$\alpha$ = Cronbach's Alpha Coefficient

$\boldsymbol{k}$ = No. of items (i.e. no. of driving performance parameters)

$\boldsymbol{S_i}^{\mathbf{2}}$ = Variance of i$^{th}$ item

$\boldsymbol{S_T}^{\mathbf{2}}$ = Variance of total score formed by summing all the items.

Before calculating Cronbach's Alpha, the items (driving performance parameters) are first multiplied by their respective weighting coefficients before calculating $\boldsymbol{S_i}^{\mathbf{2}}$ and $\boldsymbol{S_T}^{\mathbf{2}}$. The coefficient Alpha partitions the total variance among the items into signal and noise components (DeVellis, 2003) i.e., alpha is the proportion of total variation that is signal. Equation 5.1 can be interpreted as:

$$\alpha = \frac{\boldsymbol{k}}{\boldsymbol{k}-\mathbf{1}}\left(\mathbf{1}-\frac{\textbf{\textit{Sum of item variances}}}{\textbf{\textit{Sum of variances and covariances}}}\right) \tag{5.2}$$

Where the *Sum of item variances* is the sum of the variances of the items and the *Sum of variances and covariances* is the sum of all the variances and covariances of the variance-covariance matrix. Cronbach's Alpha equals the proportion of total variance among the items that is due to the underlying construct and is therefore communal (common) (DeVellis, 2003). The term on the right hand side in the brackets of Equation 5.2 is the proportion of total variance of the items that is unique; this is subtracted from 1 to get the proportion of variance that is communal (i.e. common). The term (k/k-1) is included as a correction factor in Equation 5.1 and 5.2 to ensure that an Alpha of 1 is obtained if perfect correlation is present between the items. If the items are uncorrelated, the variance of the total score (i.e., composite of the items) will be equal to the sum of the variances of individual items making the scale and so Alpha would be zero; the variance of the total score will increase as items become more and more correlated. An alternate form of Cronbach's Alpha is:

$$\alpha = \frac{\boldsymbol{k\bar{r}}}{\mathbf{1}+(\boldsymbol{k}-\mathbf{1})\boldsymbol{\bar{r}}} \tag{5.3}$$

Where, $\boldsymbol{k}$ is number of items (i.e., driving performance parameters) and $\boldsymbol{\bar{r}}$ is the average correlation between the items. Equation 5.3 shows that if Alpha is very low, the test has few items (i.e. is short) or the items have very little in common (i.e., the average correlation between items is low). According to Nunnally (1978), Alpha should be above 0.70 to demonstrate internal consistency. Streiner and Norman (2003) suggest a value greater than 0.8.

## 5.3 The Concept of Weights

Cronbach's Alpha was first estimated using unit nominal weights i.e. weighting all items equally. However, before estimating Alpha, the 24 driving performance parameters were standardized i.e., the mean of each driving performance parameter was subtracted from its respective value and then divided by the sample standard deviation of the parameter, resulting in each parameter having a mean of zero and a standard deviation of 1 (Nunnally & Bernstein, 1994). Values larger than the mean appear as positive and smaller as negative. Since standardization is a linear transformation, the distribution of each driving performance parameter was preserved (because the proportionality of inter-score distance was preserved, therefore there was no distortion in shape of distribution and correlations were not affected). Standardization does not "normalizes" (normal distribution) a variable. Standardization was necessary because the variances of the driving performance parameters were different; adding the raw scores (without standardization) would have resulted in heavier weighting of the parameters having greater variance in an apparently equally-weighed composite. It was also necessary to standardize due to the "incomparable / incompatible" nature of the parameters and/or their units of measurement (e.g., one parameter measures the no. of accidents while another measures the deviation of the vehicle from the center of the lane).

It may be noted that there is a real difference between the apparent (nominal) weight assigned to an item and the effective contribution that this weight makes to the composite or the total score, because in addition to the apparent weights, the contribution from the different items to the total score is a function of the variances of the items and the intercorrelations between the items (Helmstadter, 1964). Therefore, use of standardization (making variances equal) and then using unit nominal weights with the items comes closest to achieving equal effective weighting, particularly if the correlation of each item with the other is nearly the same (Wang & Stanley, 1970). Also in assigning nominal (apparent) weights, it is the relative magnitude of the apparent (nominal) weights of items to one another that is important rather than the absolute magnitude of such apparent weights.

## 5.4 Item Analysis and Estimating Cronbach's Alpha

### 5.4.1 Overview

The number of observations (drivers) used in the estimation of Cronbach's Alpha was 56, with each observation having 24 different driving performance parameters (items). First, all driving performance parameters were standardized. Before carrying out further estimation, it was necessary to orient (i.e., change polarity) 19 of the 24 standardized driving performance parameters so that higher scores corresponded to better driving performance (i.e., high level of the score on the item should represent high level of the construct and vice versa) (Acock, 2006; Spector, 1992; Wainer, 1976). This was accomplished by multiplying the standardized values of all 24 parameters except the following five by -1: *Time-To-Line Crossing (Mountain Drive), Coherence, No. of Correct DA Responses, Time-To-Line Crossing DA Task, Time-To-Line Crossing Car-Following*.

The standardized (and polarity corrected) driving performance parameters were multiplied with unit nominal (apparent) weights and then added up to obtain a total score (composite score). Equation 5.1 and the Pearson Moment Correlation coefficient were employed to carry out Item Analysis and estimate Alpha. Item Analysis was carried out to eliminate those items that do not contribute towards the formation of an internally consistent scale and thus lower reliability. First, the variances of the driving performance parameters and the total score (composite score) were calculated and then Equation 5.1 was used to estimate Cronbach's Alpha. Next, the Pearson Moment Correlation coefficients were found between each driving performance parameter and the total score (composite score) and also between each driving performance parameter and the total score on all the other items. This information formed the basis for the decision making process whereby certain parameters were dropped and their impact on the value of Cronbach's Alpha reliability coefficient assessed (Acock, 2006; Spector, 1992; Wainer, 1976; Kline, 1993). There is no constraint on the size of the sample relevant to the number of items (parameters) being used in Item Analysis but that it should be representative (Kline, 1993).

### 5.4.2 Three different Kind of Indices.

Three different kinds of unit nominal weight indices were computed. These were:

1. Index obtained by considering all 24 driving performance parameters.
2. Index obtained by considering all 24 driving performance parameters except *No. of Total Hazards*.
3. Index obtained by considering all 24 driving performance parameters except *No. of Total Hazards* and *No. of Low-Speed Warnings.*

The rationale for considering the three driving performance indices was that categorization of drivers with regard to driving performance was to be based on driving performance indices. To maintain the psychometric principles (described in Section 4.3.1) used in the design of the drive and the driving behaviour exhibited by drivers in this context, it was necessary to isolate the effects of *No. of Total Hazards* and *No. of Low-Speed Warnings* from the rest of the parameters. They were considered separately along with their respective indices (indices in which these parameters were not present). This was important because, as pointed out earlier, driving speed can be used to regulate the difficulty of the driving task (Hakamies-Blomqvist, 1993). Slower driving speeds have been exhibited by Alzheimer patients (compared to age matched controls) in driving simulators (Cox et al., 1998 cited in Brown & Ott, 2004) and by mildly demented subjects in on-road tests (Hunt et al., 1993 cited in Brown & Ott, 2004). Slow driving as a compensatory measure is also reported by Hakamies-Blomqvist (Hakamies-Blomqvist, 1994) and the significant difference between the speeds of older and younger drivers has been identified by McGwin & Brown (1999) as a cause of dangerous accidents. McKnight and Urquijo (1993) highlighted slow driving for identification of deficient drivers based on a sample of 1000 referral forms used by police. In simulator studies, slow speed driving by older drivers has been reported by Quillian et al. (1999), Park et al. (2007) and Chaparro and Alton (2000). Using a driving simulator, Cox et al. (1998) and Szlyk et al. (2002) report slow speed driving by Alzheimer disease subjects (out patients) and suspected-dementia subjects respectively, compared to control groups. According to Hunt et al. (1997a), results from their road test were consistent with those of other studies using road tests and showed that DAT (dementia of the Alzheimer type) subjects adopted slower speeds compared to healthy elderly controls. Staplin (Staplin et al., 1999) has classified overcautiousness (e.g., driving slowly) as a discriminating error. Discriminating errors are potentially dangerous errors that signify reduced driving skill.

### 5.4.3 Methodology

The estimation of driving performance index by consideration of all 24 driving performance parameters (Index named = *DPI1*) was undertaken as follows:

Firstly, all 24 parameters were standardized and oriented in the proper direction (by changing signs of the 19 parameters as explained in Section 5.4.1) and were added to obtain a total score. Variances of the driving performance parameters and the total score (composite score) were calculated and then Equation 5.1 was used to estimate Cronbach's Alpha. The Item Analysis and Cronbach's Alpha estimations may be seen in Table 5.1. The Bottom of Column 5 and 6 shows the average of the correlation between the 24 items and Cronbach's Alpha (based on 24 parameters) (i.e. 0.2807 and 0.9035 respectively). Column 3 gives the correlation coefficients of the driving performance parameters with the total score and column 4 gives the correlation coefficient of the parameter with the total score excluding the parameter itself. Column 5 and 6 give the average correlation coefficients between the parameters and coefficient Alpha respectively, if this particular parameter was to be removed from the analysis. Since the objective of Item Analysis is to identify weak items (weak driving performance parameters) and eliminate them so as to increase the homogeneity of the scale and thereby increase Alpha, the Item-rest correlation (column 4) was used as the criterion to make this decision (Spector, 1992; Kline, 1993) along with the value of Alpha. Nunnally and Bernstein (1994) and Kline (1993) suggest using values of Item-rest correlation (Colum 4) rather than Item-test correlation (Column 3) in order to avoid bias, as the total score includes the contribution of the item when calculating Item-test correlation (column 3). The parameter to drop first was the one whose omission had the least negative or most positive effect on Alpha; the Item-rest correlation was used as a guide in identifying the parameters which were expendable. As the lowest value of Item-rest correlation was that of parameter V31 (i.e.,-0.0387), it was dropped and the whole analysis was repeated. It may be seen from Table 5.1 that dropping V31 increases the Average inter-item correlation from 0.2807 to 0.3083 and Alpha from 0.9035 to 0.9111. For well-fitting items, Alpha will decrease if we remove the item as is evident from column 6 of Table 5.1. The table showing the rest of the analysis after dropping V31 is given in Appendix-D. The next lowest value of Item-rest correlation was for parameter V29 (i.e., 0.0061). Therefore, parameter V29 was dropped and the whole analysis repeated. This procedure was continued until the lowest value of Item-rest correlation was not less than 0.3 (Kline, 1993), at which stage the process of elimination of parameters was stopped. Tables showing details of calculations are at Appendix-D.

**Table 5.1** Item Analysis and estimation of Cronbach's Alpha

| Col.1 | Col.2 | Col.3 | Col.4 | Col.5 | Col.6 |
|---|---|---|---|---|---|
| No. | Item (Driving Performance Parameters) | Item-test correlation | Item-rest correlation | Average inter-item correlation | Alpha |
| 1 | *V25=t-hazards* | 0.7659 | 0.7322 | 0.2697 | 0.8947 |
| 2 | *V26=no-low-speedwarnings* | 0.4556 | 0.3933 | 0.2861 | 0.9021 |
| 3 | *V27=over-speed-time* | 0.7090 | 0.6687 | 0.2727 | 0.8961 |
| 4 | *V28=out-lane-time* | 0.6534 | 0.6073 | 0.2757 | 0.8975 |
| 5 | *V29=sr-mount* | 0.0841 | 0.0094 | 0.3058 | 0.9102 |
| 6 | *V30=tlc-mount* | 0.2720 | 0.2008 | 0.2958 | 0.9062 |
| 7 | *V31=abs-diff-modulus* | 0.0361 | -0.0387 | 0.3083 | 0.9111 |
| 8 | *V32=delay-phase-shift* | 0.4266 | 0.3625 | 0.2877 | 0.9028 |
| 9 | *V33=coherence* | 0.6357 | 0.5879 | 0.2766 | 0.8979 |
| 10 | *V34=no-correct-da* | 0.5873 | 0.5350 | 0.2792 | 0.8991 |
| 11 | *V35=no-da-noresponse* | 0.6118 | 0.5617 | 0.2779 | 0.8985 |
| 12 | *V36=rt-da* | 0.8215 | 0.7948 | 0.2668 | 0.8933 |
| 13 | *V37=sd-rt-da* | 0.7941 | 0.7638 | 0.2682 | 0.8940 |
| 14 | *V38=rt-stop* | 0.7077 | 0.6672 | 0.2728 | 0.8961 |
| 15 | *V39=abs-diff-speed-da* | 0.5530 | 0.4978 | 0.2810 | 0.8999 |
| 16 | *V40=sd-speed-da* | 0.7643 | 0.7304 | 0.2698 | 0.8947 |
| 17 | *V41=abs-diff-lane-pos-da* | 0.3244 | 0.2551 | 0.2931 | 0.9051 |
| 18 | *V42=sd-lane-pos-da* | 0.7315 | 0.6938 | 0.2715 | 0.8955 |
| 19 | *V43=abs-diff-lane-pos-car* | 0.3079 | 0.2380 | 0.2939 | 0.9054 |
| 20 | *V44=sd-lane-pos-car* | 0.6899 | 0.6475 | 0.2737 | 0.8966 |
| 21 | *V45=sr-da* | 0.5348 | 0.4781 | 0.2819 | 0.9003 |
| 22 | *V46=sr-car* | 0.3961 | 0.3303 | 0.2893 | 0.9035 |
| 23 | *V47=tlc-da* | 0.8327 | 0.8075 | 0.2662 | 0.8930 |
| 24 | *V48=tlc-car* | 0.6831 | 0.6400 | 0.2741 | 0.8967 |
| | | | | | |
| | Test scale | | | 0.2807 | 0.9035 |

Where,

*V25 = No. of Total Hazards*

*V26 = No. of Low-Speed Warnings*

*V27 = Over Speed Limit (Percent of Time)*

*V28 = Out of Lane (Percent of Time)*

*V29 = Steering Reversal rate (Mountain Drive)*

*V30 = Time-To-Line Crossing (Mountain Drive)*

*V31 = Absolute Difference in Modulus*

*V32 = Delay (Phase Shift)*

*V33 = Coherence*

*V34 = No. of Correct DA Responses*

*V35 = No. of DAs with No Response*

*V36 = Reaction Time DA Task*

*V37 = Standard Deviation of Reaction Time*

*V38 = Reaction Time to Stop Sign*

*V39 = Absolute Difference in Speed DA Task*

*V40 = Standard Deviation in Speed DA Task*

*V41 = Absolute Difference Lane Position DA Task*

*V42 = Standard Deviation Lane Position DA Task*

*V43= Absolute Difference Lane position Car-Following Task*

*V44 = Standard Deviation Lane position Car-Following Task*

*V45 = Steering Reversal Rate DA Task*

*V46 = Steering Reversal Rate Car-Following*

*V47 = Time-To-Line Crossing DA Task*

*V48 = Time-To-Line Crossing Car-Following*

The following five driving performance parameters were sequentially dropped which improved Alpha from 0.9035 to 0.9280: *V31*, *V29*, *V30*, *V43* and *V41*.

The same procedure was followed for the case where the index was obtained by considering all 24 driving performance parameters except *No. of Total Hazard* (Index named= *DPI2*). The initial Item Analysis was started with 23 parameters (details of the analysis are in Appendix-D). The following five driving performance parameters were sequentially dropped which improved Alpha from 0.8947 to 0.9214: *V31*, *V29*, *V41*, *V43* and *V30*.

A similar procedure was followed for the case where the index was obtained by considering all 24 driving performance parameters except *No. of Total Hazards* and *No.*

*of Low-Speed Warnings* (Index named= *DPI3*). The initial Item Analysis was started with 22 parameters (see Appendix-D). The following five driving performance parameters were sequentially dropped which improved Alpha from 0.8929 to 0.9213: *V31*, *V29*, *V41*, *V43* and *V30*. Table 5.2 summarizes the particular driving performance parameters used in forming different driving performance indices using Item Analysis and Alfa. In all three cases, high values of the reliability coefficients were obtained.

**Table 5.2** Summary of driving performance parameters included in the three driving performance indices.

| Index | Parameters Included in Index | Parameters Dropped (due to weak correlations) | Cronbach's Alpha |
|---|---|---|---|
| *DPI1* | *V25,V26,V27,V28,V32,V33,V34,V35,V36, V37,V38,V39,V40,V42,V44,V45,V46,V47,V48* | *V31,V29,V41, V43,V30* | 0.9280 |
| *DPI2* | *V26,V27,V28,V32,V33,V34,V35,V36,V37, V38,V39,V40,V42,V44,V45,V46,V47,V48* | *V31,V29,V41, V43,V30* | 0.9214 |
| *DPI3* | *V27,V28,V32,V33,V34,V35,V36,V37,V38, V39,V40,,V42,V44,V45,V46,V47,V48* | *V31,V29,V41, V43,V30* | 0.9213 |

Where,

*DPI1* = Index obtained by consideration of all 24 driving performance parameters.

*DPI2* = Index obtained by considering all 24 driving performance parameters except *No. of Total Hazard.*

*DPI3* = Index obtained by considering all 24 driving performance parameters except *No. of Total Hazards* and *No. of Low-Speed Warnings*.

The driving performance indices shown in Table 5.2 were formed by adding their respective parameters (i.e., equivalent to using unit nominal weights) under "Parameters Included in Index" after these parameters had been standardized and oriented in the proper direction (by changing signs of the 19 parameters as explained in Section 5.4.1, if any of these 19 parameters was present).

## 5.5 Weighting

### 5.5.1 Overview

The indices *DPI1*, *DPI2* and *DPI3* were computed using unit nominal weights. However, it was decided also to estimate another set of indices using differential weights based on the parameters selected in Table 5.2 for each of the three indices. These weighted versions of *DPI1*, *DPI2* and *DPI3* were called *DPI1-weighted*, *DPI2-weighted* and *DPI3-weighted* respectively. The predictive ability of an index can be increased by the use of differential weights (Perloff & Persons, 1988). According to Streiner and Norman (2003), if the items (parameters) in a scale are less than 40, differential weighting may have some effect. It is best to employ multiple regression in order to maximize the correlation between the composite score and external criterion in order to arrive at differential weights (the weights being the beta coefficients), if a reliable external criterion is present. However, in the absence of an external criterion (as in our case), a reasonable strategy would be to use differential weights to maximize the reliability of the composite (Helmstadter, 1964). For *congeneric* items, it is sensible to us differential weighting to maximize reliability (i.e.., Alpha) (Li et al., 1996). A scale is called *congeneric* when all the items are measuring the same construct i.e., when all the items are correlated with each other (Streiner & Norman, 2008). The use of differential weights to maximize reliability (Alpha) has been advocated by other authors as well (Berge & Hofstee, 1999; Nunnally & Bernstein, 1994; Wang & Stanley, 1970). More reliability translates to more variance of the composite score (i.e., more dispersion of composite score) and therefore yields more discrimination among individuals (Kline, 1993; Horst, 1966).

### 5.5.2 Weighting by Principal Component Analysis

Principal Component Analysis was used to find differential weights so as to maximize Alpha. The first Principal Component of the standardized variables (parameters) maximizes the explained variance and therefore, its eigenvector furnishes the weights that maximize Cronbach's Alpha (Lord, 1958; Berge & Hofstee, 1999). The resulting maximized Alpha is given by:

$$\alpha_{\max} = \frac{k}{k-1}\left(1 - \frac{1}{\lambda}\right) \qquad (5.4)$$

| | | |
|---|---|---|
| $\alpha_{\max}$ | = | Maximized Cronbach's Alpha. |
| $k$ | = | No. of terms (i.e. no. of parameters). |
| $\lambda$ | = | Largest Eigenvalue of the Variance-Covariance Matrix of the Standardized items (standardized parameters). |

Principal Component Analysis is a multivariate statistical method. Details of the technique can be found in Rencher (2002) and Manly (2005). In Principal Component analysis, the variance of a linear combination of the variables (parameters) is maximized. The p variables (parameters) $X_1$ , $X_2$, …., $X_p$ are taken and combinations of these are found to produce indices $Z_1$, $Z_2$, ….., $Z_p$ in order of their importance that are uncorrelated with each other and describe the variation in the data. The $Var(Z_1) \geq Var(Z_2) \geq \ldots \geq Var(Z_p)$, where $Var(Z_i)$ is the variance of $Z_i$. Since the indices are uncorrelated, that means that they are tapping different "dimensions" and are called the Principal Components. Data on the p variables (parameters) and the n observations (i.e., n drivers) are formed into a variance-covariance matrix and p Eigenvalues ( $\lambda_1, \lambda_2, ....., \lambda_p$ ) and their corresponding p Eigenvectors ( $\mathbf{a}_1$, $\mathbf{a}_2$, …, $\mathbf{a}_p$) are estimated. The first Principal Component is the linear combination of variables (parameters) $X_1$ , $X_2$, …., $X_p$ and is:

$$Z_1 = a_{11} X_1 + a_{12} X_2 + \ldots\ldots + a_{1p} X_p \qquad (5.5)$$

Where $a_{11}$, $a_{12}$, …., $a_{1p}$ are the elements of the eigenvector $\mathbf{a}_1$ corresponding to the first Eigenvalue $\lambda_1$ (which is the largest ) and so on. The Eigenvalues of the variance-covariance matrix are the variances of their corresponding Principal Components i.e. for example, variance of $Z_1$ is equal to $\lambda_1$. Also,

$$\lambda_1 + \lambda_2 + .... + \lambda_p = c_{11} + c_{22} + .... + c_{pp} \tag{5.6}$$

Where, $\lambda_1, \lambda_2, ...., \lambda_p$ are the Eigenvalues and $c_{11}$, $c_{22}$ and $c_{pp}$ are the respective variances of the variables (parameters) i.e., the diagonal elements of the variance-covariance matrix. Since the variables (parameters) were standardized (mean=0 and variance =1), therefore the sum on the right hand side of Equation 5.6 is equal to p i.e. the number of variables (parameters). The First Principal component is the linear combination that possesses the maximum variance, hence the elements of the Eigenvector corresponding to $\lambda_1$ (i.e., the largest Eigenvalue) form the weights which when used with the standardized parameters (polarity corrected) result in maximum variance (maximally separated or spread out) of the composite and hence maximum Alpha.

First, the parameters identified in Table 5.2 were standardized and oriented in the proper direction (by changing signs of the 19 parameters as explained in Section 5.4.1, if any of these 19 parameters was present). Principal component analysis was then carried out relevant to the three indices in Table 5.2 and corresponding Eigenvalues and Eigenvectors determined. Elements of the Eigenvector corresponding to the largest Eigenvalue ($\lambda_1$) constituted the weights that maximized Alpha. The maximum value of Alpha was computed using Equation 5.4 and was then verified against manual calculation using Equation 5.1. The elements of the Eigenvector corresponding to the largest Eigenvalue ($\lambda_1$) were added and each element then divided by this total to obtain relative nominal weights. The standardized parameters (polarity corrected) were then multiplied with these relative weights and added up to obtain the weighted versions of the indices *DPI1-weighted, DPI2-weighted, DPI3-weighted* corresponding to the unit nominal weight indices *DPI1*, *DPI2* and *DPI3* respectively. Table 5.3 shows the optimized weights that were obtained using Principal Component Analysis in order to estimate *DPI1-weighted*; the table also shows unit weights (relative) for comparison which were calculated as 1/19 as there were 19 parameters used in *DPI1*. The largest Eigenvalue ($\lambda_1$) was 8.60031, which was used in Equation 5.4 to estimate maximized Alpha of 0.9328 (using $k$=19). Optimized weights obtained using Principal Component Analysis for estimation of *DPI2-weighted* and *DPI3-weighted* are in Appendix-D along with values of Largest Eigenvalues and maximized Alpha. Table 5.4 summarizes Cronbach's Alpha (using unit nominal weights) for *DPI1*, *DPI2* and *DPI3* and maximized Alpha (using Principal Component weights) for *DPI1-weighted*, *DPI2-weighted* and *DPI3-weighted*. Table 5.5 shows the correlation coefficients between the six driving performance indices. It may be noted from Table 5.4 that, there was a slight increase in Alpha in all three indices when Principal Components weights were used and a very high degree of correlation between the indices. However, since differential weights change the distribution of an index and can make a difference with regard to the groupings of drivers obtained through cluster analysis, it was considered worthwhile to explore this strategy.

**Table 5.3** Optimized weights obtained using Principal Component Analysis for estimation of *DPI1-weighted* (Unit relative weights relevant to *DPI1* shown for comparison).

| No. | Driving Performance Parameters | Unit Weights (relative) | Optimized Weights (relative) |
|---|---|---|---|
| 1 | *V25=t-hazards* | 0.0526 | 0.0637 |
| 2 | *V26=no-low-speedwarnings* | 0.0526 | 0.0377 |
| 3 | *V27=over-speed-time* | 0.0526 | 0.0559 |
| 4 | *V28=out-lane-time* | 0.0526 | 0.0534 |
| 5 | *V32=delay-phase-shift* | 0.0526 | 0.0394 |
| 6 | *V33=coherence* | 0.0526 | 0.053 |
| 7 | *V34=no-correct-da* | 0.0526 | 0.0509 |
| 8 | *V35=no-da-noresponse* | 0.0526 | 0.0499 |
| 9 | *V36=rt-da* | 0.0526 | 0.0656 |
| 10 | *V37=sd-rt-da* | 0.0526 | 0.0645 |
| 11 | *V38=rt-stop* | 0.0526 | 0.0577 |
| 12 | *V39=abs-diff-speed-da* | 0.0526 | 0.0438 |
| 13 | *V40=sd-speed-da* | 0.0526 | 0.0603 |
| 14 | *V42=sd-lane-pos-da* | 0.0526 | 0.0637 |
| 15 | *V44=sd-lane-pos-car* | 0.0526 | 0.0586 |
| 16 | *V45=sr-da* | 0.0526 | 0.0379 |
| 17 | *V46=sr-car* | 0.0526 | 0.0256 |
| 18 | *V47=tlc-da* | 0.0526 | 0.0673 |
| 19 | *V48=tlc-car* | 0.0526 | 0.0511 |
| | | | |
| | Total | 1 | 1 |
| | | | |
| | Cronbach's Alpha | 0.9280 | 0.9328 |

**Table 5.4** Summary of Cronbach's Alpha using nominal unit weights for *DPI1*, *DPI2* and *DPI3* and Maximized Alpha using Principal Component weights for *DPI1-weighted*, *DPI2-weighted* and *DPI3-weighted.*

| Index | Cronbach's Alpha | Index | Maximized Cronbach's Alpha |
|---|---|---|---|
| *DPI1* | 0.9280 | *DPI1-weighted* | 0.9328 |
| *DPI2* | 0.9214 | *DPI2-weighted* | 0.9264 |
| *DPI3* | 0.9213 | *DPI3-weighted* | 0.9261 |

**Table 5.5** Correlation coefficients between the six driving performance indices.

| | DPI1 | DPI2 | DPI3 | DPI1-weighted | DPI2-weighted | DPI3-weighted |
|---|---|---|---|---|---|---|
| DPI1 | 1 | | | | | |
| DPI2 | 0.9986 | 1 | | | | |
| DPI3 | 0.9960 | 0.9971 | 1 | | | |
| DPI1-weighted | 0.9975 | 0.9944 | 0.9937 | 1 | | |
| DPI2-weighted | 0.9973 | 0.9977 | 0.9965 | 0.9980 | 1 | |
| DPI3-weighted | 0.9945 | 0.9944 | 0.9976 | 0.9968 | 0.9984 | 1 |

### 5.5.3 Weighting by Utilizing an Optimization Tool

Besides the optimized weights obtained through Principal Component analysis, it was also decided to use an optimization tool to find differential weights that maximized Cronbach's Alpha to see if a different set of weights were feasible. First, the parameters identified in Table 5.2 were standardized and oriented in the proper direction (by changing signs of the 19 parameters as explained in Section 5.4.1, if any of these 19 parameters was present). A program was devised in Microsoft Excel® in which nominal unit weights were used as starting values before being multiplied with their respective parameters, with Cronbach's Alpha being computed using Equation 5.1. The optimization tool *Solver* in Microsoft Excel® was then used to carry out a number of iterations by changing the unit nominal weights so as to maximize Alpha, with the constraints that the sum of the differential weights should add up to 1 and all differential weights should be positive. *Solver* uses the *"Generalized Reduced Gradient (GRG2) Algorithm for optimizing nonlinear problems"*. This algorithm was developed by Leon Lasdon, of the University of Texas at Austin, and Allan Waren, of Cleveland State University. Linear and integer problems use the simplex method with bounds on the variables, and the branch-and-bound method, implemented by John Watson and Dan Fylstra, Frontline Systems, Inc (http://www.solver.com/) (Microsoft Excel® Help). The optimized weights using this optimization tool were exactly the same as those obtained by using Principal Component Analysis in Section 5.5.2 and there was no change in the maximized Alpha coefficient.

## 5.6 Summary

This chapter has described the development of six different driving performance indices to be used in assessment of driving performance and the rationale for their development. The estimation of single administration reliability using Cronbach's Alpha has been emphasized along with computation formulae. The concept of weights and standardization of items (parameters) has been discussed and highlighted that the use of standardization (making variances equal) along with utilization of unit nominal weights comes closest to achieving equal effective weighting , particularly if the correlation of each item with the other is nearly the same. The procedure of Item Analysis and estimation of Cronbach's Alpha has been described. The rationale for consideration of the three different driving performance nominal-unit-weight indices (*DPI1*, *DPI2* and *DPI3*) has been elaborated upon.

The methodology used in estimating Cronbach's Alpha through Item Analysis for the three indices has been described which resulted in high values of Alpha reliability coefficients (0.9280, 0.9214 and 0.9213). Calculation of the three unit nominal weight indices has been described. The case of differential weighting has been justified and Principal Component analysis employed to find optimum weights for the weighted versions (i.e., *DPI1-weighted*, *DPI2-weighted* and *DPI3-weighted*) of the indices.

Calculation of the three weighted versions of the indices has been outlined. Based on certain driving performance parameters and driving performance indices, the performance-based categorization of drivers using a clustering technique is described in the next chapter.

# 6 Cluster Analysis

## 6.1 Introduction

The objectives of this chapter were to come up with a categorizations of drivers (based on driving performance) using the technique of normal-mixture-model cluster analysis using six different alternatives of driving performance parameter combinations/single parameters and finally the identification of the most clinically relevant grouping of drivers based on relevant driving performance parameters. After the six driving performance indices had been estimated, the next step was to make a driver-categorization based on the measures of driving performance i.e., driving performance indices and two of driving performance parameters (*No. of Total Hazards* and *No. of Low-speed Warnings*). For example, DPI2 and *No. of Total Hazards* were considered as two separate variables because they were to be used in cluster analysis to categorize/group the 56 drivers so that drivers within each group were similar with respect to these two variables and the groups were dissimilar to each other with respect to these two variables. If we did not separate *No. of Total Hazards* from DPI2 and still used cluster analysis using these two variables then the results would have been biased as the effect of *No. of Total Hazards* would be present in both variables and they would not be providing "mutually exclusive" measures of driving performance. A similar analogy goes for DPI3, *No. of Total Hazards* and *No. of Low-Speed warnings* except that now these three separate variables were to be used in cluster analysis to categorize/group the 56 drivers so that drivers within each group were similar with respect to these three variables and the groups were dissimilar to each other with respect to these three variables. The rationale for inclusion of these two driving performance parameters (*No. of Total Hazards* and *No. of Low-Speed warnings*) separately in cluster analysis has already been explained in Sections 4.4.2, 4.5.2.1 and 5.4.2. Since it is generally impossible *a priori* to perceive what combination of variables is likely to result in interesting and informative classifications (Everitt & Hothorn, 2006) of drivers, the following six possible alternatives were explored for driver-classification:

1. *DPI1*
2. *DPI1-weighted*
3. *DPI2* and *No. of Total Hazards*
4. *DPI2-weighted* and *No. of Total Hazards*
5. *DPI3*, *No. of Total Hazards* and *No. of Low-speed Warnings*
6. *DPI3-weighted*, *No. of Total Hazards* and *No. of Low-speed Warnings*

The inclusion of the number of parameters/variables in each of these six scenarios also ensured that only a small number of variables relevant to the classification were being used and that these variables described the objects/observations (Gordon, 1999; Everitt et al., 2001). It may be highlighted that *DPI1* included the effect of 19 parameters, *DPI2* and *DPI3* included the effect of the 19 parameters except "*No. of Total Hazards"* and "*No. of Total Hazards* and *Low-speed Warnings"* respectively. The "-weighted" suffix depicted their corresponding weighted versions (i.e. Principal Component weights). Previous studies (e.g., Lew et al., 2005; Schultheis et al., 2003) have categorized drivers as having "failed" when their driving performance score (as gauged by some kind of an index/score) fell more than 2 standard deviations below the mean of the control group (normal group). Assuming that a higher score on an index corresponds to better driving performance, this criterion ensures that, in a normally distributed population, approximately 2.3 percent (probability $Z \leq -2$ is 0.023) of the normal drivers will always be considered as abnormal even if their performance does not qualify for a "failed" status. Besides, this criterion cannot be used for categorization if more than one measure is used simultaneously to assess driver performance. In four of the six alternative assessment strategies highlighted above, more than one measure was considered for driving-performance assessment. Therefore, it was decided to use the statistical technique of Cluster Analysis for the categorization of drivers.

## 6.2 Clustering Methods

Details of Cluster Analysis can be found in Rencher (2002), Manly (2005), Gordon (1999), Romesburg (1984), Everitt et al. (2001), Everitt & Dunn (2001) and Everitt & Hothorn (2006). In cluster analysis, there is no *a priori* information about the underlying groupings in the data. The intent is to find out patterns in the data so as to group the multivariate observations. The groupings should be such that the objects/observations within each cluster/group are similar and the clusters/groups are dissimilar to each other. The number of groups/clusters is not usually known before hand, and the groupings should make sense in the context of the objectives/research. In many techniques of Cluster Analysis, the groupings are based on some measure of similarity/dissimilarity between all pairs of observations and/or on the maximization of some objective function. A commonly used measure of dissimilarity is the "Euclidean" distance, which is the distance between two observation vectors. Other measures of similarity/dissimilarity used are for example, "Minkowski" distance, "City Block" distance, "Canberra" distance, "Pearson Correlation", "Taxonomic" distance and "Angular Separation" etc. The use of a particular similarity / dissimilarity measure partly depends upon the type of variables involved (e.g., quantitative, qualitative, nominal, dichotomous, categorical, ordinal) and the clustering method.

Hierarchical and Partition clustering are two common approaches used to cluster observation vectors. In Hierarchical clustering, each observation is initially considered as a cluster, the two groups that are closest are combined (n-1 groups, one of size two and the rest of size1), and this process continues till a single cluster containing all n observations is obtained. A hierarchy of clusters is generated by this process. This process can also be applied in reverse manner i.e. initially a single cluster containing all n observations is considered and finally n clusters having one observation each is obtained. For example, the Single Linkage, Complete Linkage, Average Linkage, Centroid and Ward's Method fall under the rubric of Hierarchical clustering. In Partition clustering, initial partitioning is used to divide the observations into g clusters or g cluster centers are initially demarcated; then some optimality criterion forms the basis of reallocating observations between / to these g clusters. *K-means* clustering comes under the rubric of Partition clustering. There are Model-based clustering methods that employ finite mixture densities as models for cluster analysis; this entails estimation of parameters of the assumed mixture and then the posterior probabilities of cluster membership.

In the Hierarchical methods, the process is irreversible i.e., once two observations / clusters are merged/separated into clusters, they cannot be separated / merged later in the procedure so mistakes cannot be corrected (Rencher, 2002). Also, consideration needs to be given to whether the hierarchical structure that is imposed on the data is acceptable or it introduces unacceptable distortion of the original relationships amongst the observations relevant to their similarity/dissimilarity measures (Everitt et al., 2001). Even the results from an optimization method such as the *K-means* can be substantially affected by the choice of the initial partition (Everitt et al., 2001); with data that is not well structured, different initial partitions may lead to different *local* optima of the clustering criteria rather than a *global* optimum. Also, there is the controversial issue of standardization (i.e. subtracting the mean of each variable from its value and then dividing by the Standard deviation of the variable so that the means of all variables are zero and their variances are all one), which is essentially equalizing their variances. Standardization can greatly affect the groupings that are discovered (Stata, 2007); standardization is sometimes necessary in order to prevent a variable having high variability from dominating a cluster analysis, while as in other cases it can hide the true grouping inherent in the data because if a particular variable separates the groups well, then the variance of this variable will be large but standardization will tend to equalize it. Therefore, the question of whether to standardize or not is not an easy one (Rencher, 2002).

The commonly used Hierarchical and Partition clustering methods, although based on intuitively reasonable procedures (Everitt, 2005) are heuristic in nature, as the allocation of objects to clusters is not based on rules that have an underlying statistical model (probabilistic or deterministic) but rather clustering is brought about through optimization of an objective function which is based on bringing about cohesion within clusters and isolation between clusters (Leese & Landau, 2006). The hieuristic methods treat variables as independent within clusters. Also, different similarity / dissimilarity measures and clustering algorithms give widely different cluster groupings and there is no solid ground on which to prefer one method over the other. Generally the statistical properties of the heuristic methods are not known which excludes the possibility of formal inference (Fraley & Raftery, 2002; Everitt & Hothorn, 2006). Furthermore, there is scant systematic guidance for addressing the fundamental and vital issue of ascertainment / determination of the number of clusters (Fraley & Raftery, 2002; Everitt & Hothorn, 2006). Model based clustering is preferred in this respect (Leese & Landau, 2006) and is increasing superseding the older heuristic methods (Everitt, 2005). The issue of standardization in model-based (which is based on distributions) clustering is not an impediment as clustering using standardized (which is a linear transformation and does not distorts the distribution of a variable) and un-standardized variables gives identical clusters. Variables that are highly correlated within groups/clusters lead to clusters that are ellipsoidal in p-dimensional space; however, the heuristic clustering methods typically fail to recognize this true structure and impose spherical structure on the data (Papageorgiou & Baxter, 2001). Physically separate clusters (i.e. mutually exclusive) are easy to distinguish with most clustering criteria / methods, however, the Hierarchical and Partition clustering methods perform poorly (compared to model-based methods) in separating true groups when there is overlap between the groups or when they intersect (Banfield & Raftery, 1993; Fraley & Raftery, 1998). The need for deciding which inter-individual similarity / dissimilarity measure to use is completely dispensed with when a model-based clustering approach is adopted (Everitt & Dunn, 1991). The model-based approach is essentially based on the idea that the data consists of independent samples from a series of group/cluster populations, but the group/cluster labels of each independent sample have been lost, so the data can be considered as coming from a mixture distribution (usually a mixture of multivariate normal distributions). The next step is to estimate the parameters of each multivariate distribution in the mixture and then the posterior probabilities of group/cluster membership of each observation/object (Venables & Ripley, 2002).

## 6.3 Model-Based Clustering / Normal Mixture Models

As highlighted by Everitt and Hothorn (2006), in model based clustering, the most successful approach was first proposed by Scott and Symons (1971), extended by Banfield and Raftery (1993), and then further extended by Fraley and Raftery (1999, 2002). The following description of the method has been primarily taken from Fraley and Raftery (1999, 2002). It is assumed that the sample of observations /objects comes from a population that consists of G subpopulations each depicting a cluster/group. The data is given by **y** which has independent multivariate observations given by $\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_n}$ , the likelihood function for a mixture model is given by

$$L_{MIX}\left(\theta_1,\ldots,\theta_G;\tau_1,\ldots,\tau_G \mid \mathbf{y}\right)=\prod_{i=1}^{n}\sum_{k=1}^{G}\tau_k f_k\left(\mathbf{y}_i \mid \theta_k\right) \tag{6.1}$$

Where $f_k$ and $\theta_k$ are the density and parameters respectively of the *k*th component in the mixture and $\tau_k$ is the probability that an observation/object belongs to the *k*th component ( $\tau_k \geq 0$; $\sum_{k=1}^{G}\tau_k = \mathbf{1}$). The multivariate normal (Gaussian) density ( $\phi_k$ ) is commonly used because of its computational tractability (McLachlan & Chang, 2004). The parameters of this multivariate normal density are $\mu_k$ which is the mean and the covariance matrix, $\Sigma_k$ . The multivariate density is

$$\phi_k\left(\mathbf{y}_i \mid \mu_k, \Sigma_k\right) \equiv \frac{\exp\left\{-\tfrac{1}{2}\left(\mathbf{y}_i-\mu_k\right)^T \Sigma_k^{-1}\left(\mathbf{y}_i-\mu_k\right)\right\}}{\sqrt{\det\left(2\pi\Sigma_k\right)}} \tag{6.2}$$

The multivariate normal mixture densities generate data that is characterized by clusters that are centered at the means $\mu_k$ . The density of points that are nearer to the mean is greater. The surfaces of constant density have an ellipsoidal shape. Other geometric properties of the clusters are determined by the variance-covariance matrix $\Sigma_k$ . Usually, it is impractical to maximize the function in Equation 6.1 without imposing some constraints on the parameters. Characteristics (shape, volume and orientation) are usually estimated from the data and can be constrained to be the same for all clusters or allowed to vary between clusters. This is brought about through the eigenvalue decomposition of the variance-covariance matrix $\Sigma_k$ of the multivariate normal distribution for the $k^{\text{th}}$ component given by

$$\Sigma_k = \lambda_k \boldsymbol{D}_k \boldsymbol{A}_k \boldsymbol{D}_k^{\boldsymbol{T}} \tag{6.3}$$

Where,

d = Dimensionality of the data

$\Sigma_k$ = Variance-covariance matrix

$\boldsymbol{D}_k$ = Orthogonal d × d matrix of eigenvectors

$\boldsymbol{A}_k$ = d × d diagonal matrix whose elements are proportional to the eigenvalues of $\Sigma_k$

$\lambda_k$ = An associated constant of proportionality (it is a scalar)

$\boldsymbol{D}_k{}^T$ = Transpose of the orthogonal matrix of eigenvectors

$\lambda_k$, $\boldsymbol{A}_k$, and $\boldsymbol{D}_k{}^T$ are considered as independent sets of parameters relevant to each cluster and either they are constrained to be the same for each cluster or are allowed to vary among clusters. When these parameters are fixed, clusters share certain geometric properties. The orientation of the $k^{\text{th}}$ component of the mixture is governed by $\boldsymbol{D}_k$ ; $\boldsymbol{A}_k$ determines the shape of density contours and $\lambda_k$ specifies the volume (size) of the relevant ellipsoid. The parameterization of the covariance matrix also includes (but is not restricted to) the well known models: (1) where all clusters are spherical and of the same size ($\Sigma_k = \lambda \boldsymbol{I}$) which gives the sum of squares criteria (Ward, 1963) (2) When the covariance matrix is equal across all components/clusters ($\Sigma_k = \Sigma$) where the geometry of all clusters is the same but not necessarily spherical (Friedman & Rubin, 1967) (3) Unrestrained $\Sigma_k$, where each cluster has a different geometry (Scott and Symons, 1971). To characterize the covariance structure of the mixture, in case when $\Sigma_k = \lambda \boldsymbol{I}$, $\Sigma_k = \Sigma$ and $\Sigma_k$ is unrestricted, one, d(d+1)/2 and G(d(d+1)/2) parameters are required respectively, where d is the dimensionality of the data and G is the number of clusters/components.

The geometric interpretation of various parameterizations and the multivariate model options that have been provided in the software package MCLUST for hierarchical clustering (HC) (here model-based) and EM (Expectation Maximization) that runs in the statistical package R® which is based on the S programming language are shown in Table 6.1 from Fraley and Raftery (2006). MCLUST was downloadable for free for academic research purposes from the University of Washington website at http://www.stat.washington.edu/mclust. Documentation for the MCLUST package has been provided by Fraley and Raftery (2006), with minor revisions incorporated in 2007.

From Table 6.1 it may be seen that when the dimensionality of the data is one (i.e., when clustering is based on only one variable i.e. univariate case), there are only two models: (1) Equal variance (designated by E in the model identifier column), and (2) varying variance (designated by V in the model identifier column). In table 6.1, for example, EVI stands for a model in which the volumes of all clusters are equal (E), the shapes may vary (V) and the orientation is the identity (I). The covariances of clusters in this model are diagonal with orientation parallel to the coordinate axes. Sample data is used to determine the parameters that are associated with characteristics signified by E or V.

**Table 6.1** Parameterization of the covariance matrix $\Sigma_k$ currently available in MCLUST for hierarchical clustering (HC) (model-based) and/or EM for multidimensional data.

| Identifier | Model | HC | EM | Distribution | Volume | Shape | Orientation |
|---|---|---|---|---|---|---|---|
| E | | • | • | (univariate) | equal | | |
| V | | • | • | (univariate) | variable | | |
| EII | $\lambda \boldsymbol{I}$ | • | • | Spherical | equal | equal | NA |
| VII | $\lambda_k \boldsymbol{I}$ | • | • | Spherical | variable | equal | NA |
| EEI | $\lambda \boldsymbol{A}$ | | • | Diagonal | equal | equal | coordinate axes |
| VEI | $\lambda_k \boldsymbol{A}$ | | • | Diagonal | variable | equal | coordinate axes |
| EVI | $\lambda \boldsymbol{A}_k$ | | • | Diagonal | equal | variable | coordinate axes |
| VVI | $\lambda_k \boldsymbol{A}_k$ | | • | Diagonal | variable | variable | coordinate axes |
| EEE | $\lambda \boldsymbol{DAD}^T$ | • | • | Ellipsoidal | equal | equal | equal |
| EEV | $\lambda \boldsymbol{D}_k \boldsymbol{A} \boldsymbol{D}_k^T$ | | • | Ellipsoidal | equal | equal | variable |
| VEV | $\lambda_k \boldsymbol{D}_k \boldsymbol{A} \boldsymbol{D}_k^T$ | | • | Ellipsoidal | variable | equal | variable |
| VVV | $\lambda_k \boldsymbol{D}_k \boldsymbol{A}_k \boldsymbol{D}_k^T$ | • | • | Ellipsoidal | variable | variable | variable |

Note: The • in the appropriate column indicates availability. (Source: Fraley and Raftery, 2006).

The methods for maximum likelihood clustering are based on the EM (Expectation-Maximization) algorithm (Dempster et al., 1977; McLachlan & Krishnan, 1997). This procedure is in fact a particular example of the EM algorithm relevant to Maximum Likelihood estimation of missing data (i.e. here the label of the component density to which an observation belongs is missing) (Everitt, 1996). In EM relevant to mixture model clustering, $\mathbf{x}_i = (\mathbf{y}_i, \mathbf{z}_i)$ are considered the "complete" data , where $\mathbf{z}_i = (z_{i1}, \ldots., z_{iG})$ is the unobserved portion of the data with

$$z_{ik} = \begin{cases} 1 & \text{if } \mathbf{x_i} \text{ belongs to group k} \\ 0 & \text{otherwise.} \end{cases}$$

Assuming that each $\mathbf{z}$ is independent and identically distributed and has a multinomial distribution with probability of one draw from G categories as $\tau_1, \ldots., \tau_G$, the resulting density function of $\boldsymbol{y}_i$ given $\boldsymbol{z}_i$ is given by $\prod_{k=1}^{G} f_k(\mathbf{y_i} \mid \theta_k)^{z_{ik}}$ and the complete-data log-likelihood is

$$l(\theta_k, \tau_k, z_{ik} \mid \mathbf{x}) = \sum_{i=1}^{n} \sum_{k=1}^{G} z_{ik} \log\left[\tau_k f_k(\mathbf{y_i} \mid \theta_k)\right] \tag{6.4}$$

An iteration of EM comprises of an "E"-step and an "M"-step. The "E"-step for mixture models is given by

$$\hat{z}_{ik} \leftarrow \frac{\hat{\tau}_k f_k\left(\mathbf{y_i} \mid \hat{\theta}_k\right)}{\sum_{j=1}^{G} \hat{\tau}_j f_j\left(\mathbf{y_i} \mid \hat{\theta}_j\right)} \tag{6.5}$$

In the "E"-step, a matrix **z** is computed such that each element $z_{ik}$ is an estimate of the conditional probability (posterior probabilities of belonging to a particular cluster) that observation *i* belongs to cluster *k* given the current estimates of the parameters. The "M"-step then computes the maximum likelihood estimates of the parameters based on these *z* values i.e., the "M"-step consists of maximizing Equation 6.4 in terms of the parameters $\tau_k$ and $\theta_k$ with $z_{ik}$ fixed at values obtained from the previous "E"-step, $\hat{z}_{ik}$. The procedure alternates between these two steps until convergence takes place. Each iteration results in improved estimates of posterior probabilities of cluster membership. The parameters usually converge to the maximum likelihood values for the normal mixture model given by Equation 6.1 and the column means of the matrix **z** converge to the mixing proportions $\tau_k$, where G is the number of groups/clusters which is fixed before implementing the EM algorithm. An initial estimate of **z** (for starting the iterations) is obtained from a discrete classification which results in matrix **z** having zero and one entries with exactly one 1 per row. This discrete classification is achieved through model-based hierarchical clustering (HC) which uses classification maximum likelihood approach to determine which two groups to merge at each stage (Banfield & Raftery, 1993; Celeux & Govaert, 1993; Fraley, 1998).

The value $z^*_{ik}$ of $\hat{z}_{ik}$ at convergence is the estimated conditional probability that of observation $i$ belonging to group $k$. The maximum likelihood classification of observation $i$ is $\left\{ j \mid z^*_{ij} = \max_k z^*_{ik} \right\}$ i.e. the maximum value in the row of matrix **z** for a particular observation. The uncertainty associated with the classification of an observation is given by $\left(1-\max_k z^*_{ik}\right)$. For multivariate normal mixtures, $f_k$ in Equation 6.5 (which is the "E"-step) is replaced by $\phi_k$ defined in Equation 6.2, regardless of the parameterization.

One significant benefit of using a mixture-model approach to cluster analysis is that it enables the comparison of models through the use of reliable approximate Bayes factors. Thus, the selection of parameterization of the model (i.e. the clustering method) and the number of clusters/groups are brought about systematically. The Bayes factor used is given by BIC (Schwarz, 1978), which is:

$$\mathrm{BIC} = 2l_M - m_M \log_e(n) \tag{6.6}$$

Where,

$l_M$ = Maximized mixture loglikelihood for the model

$m_M$ = No. of independent parameters to be estimated in the model

$n$ = No. of observations

In general, larger values of BIC correspond to stronger evidence for a model and the number of clusters. There is a concomitant increase in the loglikelihood as the number of clusters/components increase (and hence $m_M$) (McLachlan & Chang, 2004). As more terms are added to a model, the fit of the mixture model to a given data set improves (i.e., the loglikelihood increases); therefore, model assessment through the sole use of loglikelihood can not be used for model assessment in mixture-model cluster analysis.

The second term in Equation 6.6 (i.e. $m_M \log_e(n)$) is subtracted from the loglikelihood, thus penalizing the complexity of the model, so that it may be maximized for more parsimonious parameterizations and smaller number of clusters/groups (Fraley & Raftery, 1998).

The assumption that the data set have a multivariate normal distribution is not an essential requirement for mixture-model cluster analysis (Papageorgiou & Baxter, 2001; Fraley and Raftery, 2002). In mixture-models, if the EM (Expectation Maximization) algorithm for a model having a certain number of components (clusters/groups) is applied to a mixture in which in reality there are fewer number of groups/clusters, then it may fail due to ill-conditioning. Ill-conditioning occurs due to the singularity or near singularity of the covariance matrix associated with one or more components. Therefore, it is important to avoid using the procedure to find a larger number of components (clusters/groups) than is necessary (Fraley & Raftery, 1998).

## 6.4 Clustering Scenarios

Since our sample of drivers was from the general driver population (and not from a dementia clinic etc), the only *a priori* information that would assist in the identification of drivers with poor driving skills (labeled the deficient-driver group below) was:

- The deficient-driver group will be a relatively smaller group (in size).
- On average, the deficient-driver group will have unfavourable scores (compared with the other group/groups) on the variables/parameters used in the classification of drivers.
- The cluster analysis scenario in which the classification of drivers takes into account the *No. of Total Hazards* and/or the *No. of Low-speed Warnings* (along with the relevant driving performance index) will hold more promise due to reasons explained in Sections 4.3.1, 4.4.2 and 5.4.2.
- Based on subject knowledge, the meaningful/relevant total number of categories/groups identified by cluster analysis based on driving performance will be 2 or 3.

Whilst noting the above points, it was necessary to explore all six scenarios outlined in Section 6.1 and select the one that is the most appropriate for identification of deficient drivers. Normal-Mixture model based cluster analysis was performed using the statistical software R® and the add-on package MCLUST. Except for the univariate case in Sections 6.4.1 and 6.4.2, all ten candidate parameterizations in Table 6.1 were run, i.e. EII, VII, EEI, VEI, EVI, VVI, EEE, EEV, VEV and VVV with each number of clusters from 2 to 9. Although it was considered that outcomes for higher number of clusters (say greater than 3) would not be likely to add value to the understandings, the default setting in the software (of up to 9 clusters), was retained as the BIC was to be used as a guide in identifying the appropriate parameterization (model) and the number of clusters. For the univariate case, the parameterization E and V (in Table 6.1) were used with each number of clusters from 2 to 9. A matrix of BIC values corresponding to each combination of the number of clusters and parameterization (model) was obtained along with cluster membership and uncertainty of classification. It may be noted that in certain combinations of the number of clusters and parameterization, ill-conditioning occurred (denoted by NA, i.e. BIC value was not returned). This especially occurred where the number of clusters were more than in reality, and was irrelevant. The detail of the cluster analysis using the six scenarios is described in the following sections and in Appendix-E. It may be highlighted that the group numbers are arbitrary and that when groups intersect/overlap, drivers having relatively higher values of uncertainty of classification would be expected to fall in the overlapping region (Fraley and Raftery, 2006).

### 6.4.1 Cluster Analysis using *DPI1*

In this scenario, only one variable (univariate case i.e. dimensionality of the data was one) i.e. *DPI1* was used in cluster analysis, because it was the composite of all 19 driving performance parameters shown in Table 5.2. The maximum value of this index was 21.912 and the minimum -36.495. Appendix-E contains all detailed output of the analysis. The three models with the highest BIC values are shown in Table 6.2. Model E with 2 groups/clusters has the highest BIC value. The number of drivers in each group and their group-mean value on the index *DPI1* are shown in table 6.3. Clearly, on average, group no. 2 has higher scores on the index than group no. 1, which is smaller with 14 drivers. Therefore group no. 1 may be classified as having deficient driving skills. Although group no.1 is small, its size is relatively large compared with the total size of the sample. Other ancillary plots can be found in Appendix-E.

**Table 6.2** Best three models along with BIC values and number of groups/clusters for cluster analysis using *DPI1*.

| Best BIC Values | | |
|---|---|---|
| E,2 | V,2 | V,1 |
| -445.382 | -448.071 | -449.197 |

**Table 6.3** Mean values of *DPI1* for groups/clusters using model E with BIC of -445.382

| Groups for model E with 2 clusters/groups having BIC of -445.382 | | |
|---|---|---|
| Group No. | No. of Drivers | Mean *DPI1* |
| 1 | 14 | -18.211467 |
| 2 | 42 | 6.070489 |

### 6.4.2 Cluster Analysis using *DPI1-weighted*

In this scenario, only one variable (univariate case i.e. dimensionality of the data was one) i.e. *DPI1-weighted* was used in cluster analysis because it was the weighted composite of all 19 driving performance parameters identified in Table 5.2; the weights being the Principal Component weights derived in Section 5.5.2. The maximum value of this index was 1.155 and the minimum -2.129.

The three models with the highest BIC values are shown in Table 6.4. Model E with 2 groups/clusters has the highest BIC value. The number of drivers in each group and their group-mean value on the index *DPI1*-weighted are shown in Table 6.5. Clearly, on average, group no. 2 has higher scores on the index compared to group no. 1, which is a smaller group with 13 drivers. Therefore drivers in group no. 1 may be considered as possessing poor driving skills. Although group no.1 is small, its size is still relatively large compared with the total size of the sample. Other ancillary plots can be found in Appendix-E.

**Table 6.4** Best three models along with BIC values and number of groups/clusters for cluster analysis using *DPI1-weighted*.

| Best BIC Values | | |
|---|---|---|
| E,2 | V,2 | V,1 |
| -120.441 | -122.430 | -124.013 |

**Table 6.5** Mean values of *DPI1-weighted* for groups/clusters using model E with BIC of -120.441

| Groups for model E with 2 clusters/groups having BIC of -120.441 | | |
|---|---|---|
| Group No. | No. of Drivers | Mean *DPI1-weighted* |
| 1 | 13 | -1.0413267 |
| 2 | 43 | 0.3148197 |

## 6.4.3 Cluster Analysis using *DPI2* and *No. of Total Hazards*

In this scenario, two variables (dimensionality of the data was two) i.e. *DPI2* and *No. of Total Hazards* were used in cluster analysis. The variable *No. of Total Hazards* was included because *DPI2* did not include the effect of this variable. The maximum value of *DPI2* was 20.915 and the minimum -32.211. The variable *No. of Total Hazards* had a maximum value of 30 and a minimum of 2.

The results of the three models with the highest BIC values are shown in Table 6.6. Model EEV with 2 groups/clusters has the highest BIC value. The number of drivers in each group and their group-mean values on the index *DPI2* and *No. of Total Hazards* are shown in Table 6.7. On average, group no. 1 has higher scores on the index *DPI2* and lower *No. of Total Hazards* compared with group no. 2, which is a smaller group with 14 drivers. Drivers in group no. 2 may be considered as possessing poor driving skills. Although group no.2 is small, its size is still relatively large compared with the total size of the sample. Other ancillary plots can be found in Appendix-E.

**Table 6.6** Best three models along with BIC values and number of groups/clusters for cluster analysis using *DPI2* and *No. of Total Hazards*.

| Best BIC Values | | |
|---|---|---|
| EEV,2 | VEV,2 | VVV,2 |
| -736.709 | -739.345 | -743.235 |

**Table 6.7** Mean values of *DPI2* and *No. of Total Hazards* for groups/clusters using model EEV with BIC of -736.709.

| Groups for model EEV with 2 clusters/groups having BIC of -736.709 | | | |
|---|---|---|---|
| Group No. | No. of Drivers | Mean *No of Total Hazards* | Mean *DPI2* |
| 1 | 42 | 6.97619 | 5.677471 |
| 2 | 14 | 14.71429 | -17.032412 |

### 6.4.4 Cluster Analysis using *DPI2-weighted* and *No. of Total Hazards*

In this scenario, two variables (dimensionality of the data was two) i.e. *DPI2-weighted* and *No. of Total Hazards* were used in cluster analysis. The variable *No. of Total Hazards* was included because *DPI2-weighted* did not include the effect of this variable. The maximum value of *DPI2-weighted* was 1.170 and the minimum -1.959. The variable *No. of Total Hazards* had a maximum value of 30 and a minimum of 2.

The three models with the highest BIC values are given in Table 6.8. Model VEV with 2 groups/clusters has the highest BIC value. The number of drivers in each group and their group-mean values on the index *DPI2-weighted* and *No. of Total Hazards* are shown in Table 6.9. On average, group no. 1 has higher scores on the index *DPI2-weighted* and lower *No. of Total Hazards* compared with group no. 2, which is a smaller group with 16 drivers. Drivers in group no. 2 may be considered as possessing poor driving skills. Although group no.2 is small, its size is still relatively large compared with the total size of the sample. Other ancillary plots can be found in Appendix-E.

**Table 6.8** Best three models along with BIC values and number of groups/clusters for cluster analysis using *DPI2-weighted* and *No. of Total Hazards*.

| Best BIC Values | | |
|---|---|---|
| VEV,2 | EEV,2 | VVV,2 |
| -420.308 | -420.553 | -420.813 |

**Table 6.9** Mean values of *DPI2-weighted* and *No. of Total Hazards* for groups/clusters using model VEV with BIC of -420.308.

| Groups for model VEV with 2 clusters/groups having BIC of -420.308 | | | |
|---|---|---|---|
| Group No. | No. of Drivers | Mean *No of Total Hazards* | Mean *DPI2-weighted* |
| 1 | 40 | 6.55 | 0.3124784 |
| 2 | 16 | 14.8125 | -0.7811958 |

## 6.4.5 Cluster Analysis using DPI3, No. of Total Hazards and No. of Low-speed Warnings

In this scenario, three variables (dimensionality of the data was three) i.e. *DPI3, No. of Total Hazards* and *No. of Low-speed Warnings* were used in cluster analysis. The variables *No. of Total Hazards* and *No. of Low-speed Warnings* were included because *DPI3* did not include the effects of these variables. The maximum value of this index was 19.554 and the minimum -32.845. The variable *No. of Total Hazards* had a maximum value of 30 and a minimum of 2 and *No. of Low-speed Warnings* had a maximum value of 656 and a minimum of 286.

The three models with the highest BIC values are shown in table 6.10. Model VEI with 2 groups/clusters has the highest BIC value. The number of drivers in each group and their group-mean values on the index *DPI3, No. of Total Hazards* and *No. of Low-speed Warnings* are shown in Table 6.11. On average, group no. 2 has higher scores on the index *DPI3,* lower *No. of Total Hazards* and lower *No. of Low-speed Warnings* compared to group no. 1, which is a smaller group with 18 drivers. Drivers in group no. 1 may be considered as possessing poor driving skills. Although group no.1 is small, its size is still relatively large compared to the total size of the sample. Other ancillary plots can be found in Appendix-E.

**Table 6.10** Best three models along with BIC values and number of groups/clusters for cluster analysis using *DPI3, No. of Total Hazards* and *No. of Low-speed Warnings.*

| Best BIC Values | | |
|---|---|---|
| VEI,2 | VVI,2 | VEV,2 |
| -1386.271 | -1390.824 | -1391.108 |

**Table 6.11** Mean values of *DPI3, No. of Total Hazards* and *No. of Low-speed Warnings* for groups/clusters using model VEI with BIC of -1386.271.

| Groups for model VEI with 2 clusters/groups having BIC of -1386.271 | | | | |
|---|---|---|---|---|
| Group No. | No. of Drivers | Mean *No. of Total Hazards* | Mean *No. of Low Speed warnings* | Mean *DPI3* |
| 1 | 18 | 13.8333 | 500.8333 | -13.0564 |
| 2 | 38 | 6.5789 | 392.3421 | 6.1846 |

### 6.4.6 Cluster Analysis using DPI3-weighted, No. of Total Hazards and No. of Low-speed Warnings

In this scenario, three variables (dimensionality of the data was three), i.e. *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings*, were used in the cluster analysis. The variables *No. of Total Hazards* and *No. of Low-speed Warnings* were included because *DPI3-weighted* did not include the effects of these variables. The maximum value of *DPI3-weighted* index was 1.159 and the minimum -2.084. The variable *No. of Total Hazards* had a maximum value of 30 and a minimum of 2 and *No. of Low-speed Warnings* had a maximum value of 656 and a minimum value of 286.

The three models with the highest BIC values are shown in table 6.12. Model VEI with 2 groups/clusters has the highest BIC value. The number of drivers in each group and their group-mean values on the index *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* are shown in Table 6.13. On average, group no. 2 has higher scores on the index *DPI3-weighted,* lower *No. of Total Hazards* and lower *No. of Low-speed Warnings* compared to group no. 1, a smaller group with 18 drivers. Drivers in group no. 1 may be considered to possess poor driving skills. Although group no.1 is small, its size is still relatively large compared with the total size of the sample. Other ancillary plots can be found in Appendix-E. It may be noted that the grouping of drivers obtained in this scenario is exactly the same as that described in Section 6.4.5 (i.e., the same drivers were included in both groups in both scenarios). One interesting observation in this scenario was that a 3 cluster grouping was also among the best three, as evident from Table 6.12 with a small difference in BIC from the one with the highest value. Therefore, it was decided to use the three variables i.e. *DPI3-weighted, No. of Total Hazards* and *No. of*

*Low-speed Warnings* in cluster analysis by considering all ten models with the number of clusters being 3, which is described in the next section.

**Table 6.12** Best three models along with BIC values and number of groups/clusters for cluster analysis using *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings.*

| Best BIC Values | | |
|---|---|---|
| VEI,2 | EEV,3 | VEV,2 |
| -1070.690 | -1075.632 | -1076.110 |

**Table 6.13** Mean values of *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* for groups/clusters using model VEI with BIC of -1070.690.

| Groups for model VEI with 2 clusters/groups having BIC of -1070.690 | | | | |
|---|---|---|---|---|
| Group No. | No. of Drivers | Mean *No. of Total Hazards* | Mean *No. of Low Speed warnings* | Mean *DPI3-weighted* |
| 1 | 18 | 13.833333 | 500.8333 | -0.797444 |
| 2 | 38 | 6.578947 | 392.3421 | 0.3777367 |

### 6.4.7 Cluster Analysis using DPI3-weighted, No. of Total Hazards and No. of Low-speed Warnings considering only 3 clusters

In this scenario, three variables i.e. *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* were used in cluster analysis by considering all ten models with the number of clusters being 3.

The three models with the highest BIC values for 3 clusters are shown in Table 6.14. Model EEV has the highest BIC value. The number of drivers in each group and their group-mean values on the index *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* are shown in Table 6.15. On average, group no. 3 has the highest scores on the index *DPI3-weighted,* the lowest *No. of Total Hazards* and the lowest *No. of Low-speed Warnings* compared with group no. 1 and 2. Also, group no.1 has higher score on the index *DPI3-weighted,* lower *No. of Total Hazards* and lower *No. of Low-speed Warnings* compared to group no.2. Therefore, driving performance in decreasing order of skill by group number is: group no.3, no.1 and no. 2 (see Table 6.15 and Figure 6.1).

Group no.2 is the smallest group (comprising of 8 drivers) and may be considered as possessing poor driving skills. Other ancillary plots can be found in Appendix-E.

**Table 6.14** Best three models along with BIC values for 3 clusters/groups using *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings*.

| Best BIC Values | | |
|---|---|---|
| EEV,3 | VEI,3 | EEI,3 |
| -1075.632 | -1079.265 | -1083.289 |

**Table 6.15** Mean values of *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* for model EEV with BIC of -1075.632.

| Groups for model EEV with 3 clusters/groups having BIC of -1075.632 | | | | |
|---|---|---|---|---|
| Group No. | No. of Drivers | Mean *No. of Total Hazards* | Mean *No. of Low-Speed warnings* | Mean *DPI3-weighted* |
| 1 | 20 | 8.1 | 441.55 | -0.2607223 |
| 2 | 8 | 16.5 | 553 | -0.8960832 |
| 3 | 28 | 7.321429 | 381.0357 | 0.4422541 |

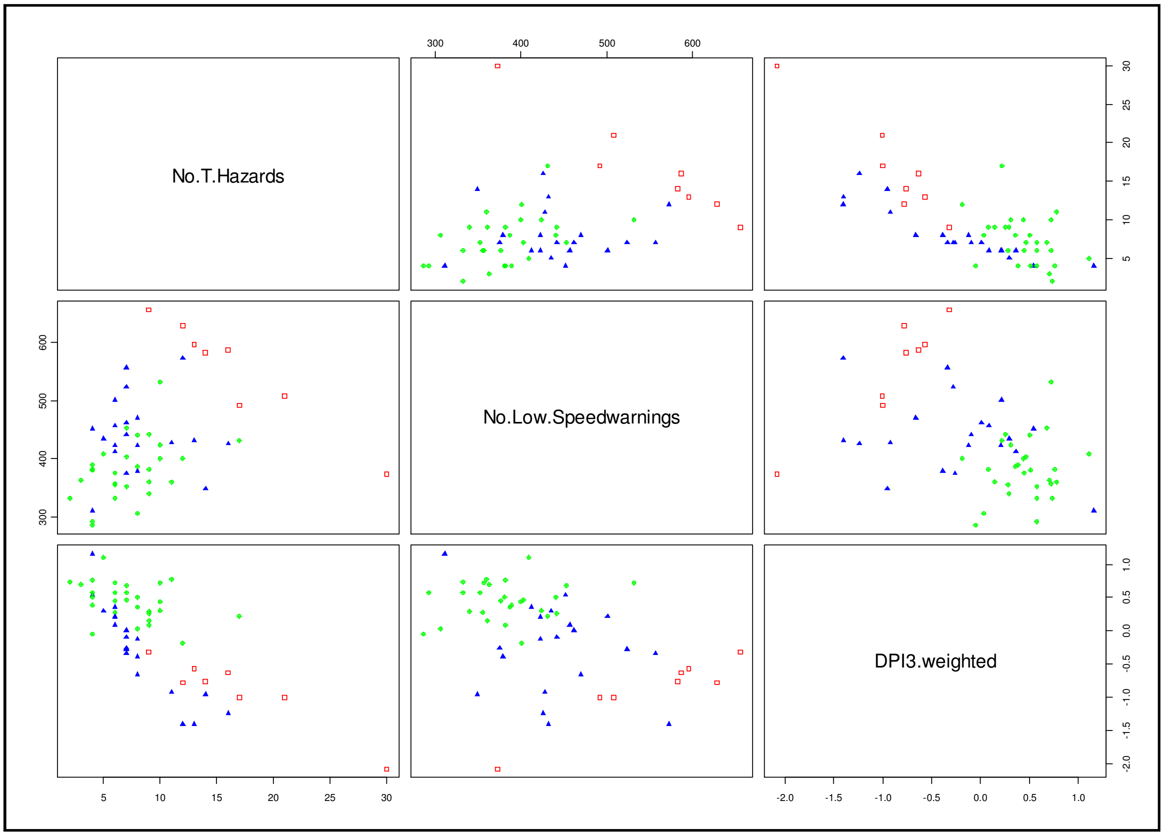



**Figure 6.1** Scatter-matrix plot of *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* showing driver classification using the model EEV with 3 groups/clusters having BIC of -1075.632. The square symbols represent group No.2, the triangular symbols represent group No.1 and the other symbols (i.e. circular symbols) represent group No.3.

### 6.4.8 K-Means Cluster Analysis using DPI3-weighted, No. of Total Hazards and No. of Low-speed Warnings.

Since the K-means clustering method is gaining more use than the hierarchical methods (Afifi et al., 2003), it was also decided to apply K-means clustering to *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* for assessment of clusters. The statistical package R® with the built-in function *kmeans* was used for the purpose.

In the K-means clustering technique, a set of data is partitioned into a specific number of groups/clusters which minimizes the within-group sum of squares (sum of squared deviations of each point from its cluster centroid on every dimension) over all variables (Hartigan & Wong, 1979). Apparently, the problems seems relatively simple in that every possible partition of the n individuals/objects/observations into *k* groups should be considered and the one with the lowest within-group sum of squares selected. However, in actual practice, the numbers involved are so vast that it is impossible to completely enumerate every possible partition even with the fastest computer (Everitt & Hothorn, 2006). This has led to the development of algorithms designed to search for configurations with the minimum within-group sum of squares by reallocating observations/objects between groups and keeping the new one only if it provides an improvement (Everitt & Hothorn, 2006). Although, such algorithms do not guarantee that the global minimum will be found. The essential steps of these algorithms are:

1. An initial partition of individuals/objects/observations is made into *k* groups either by using one of the hierarchical clustering techniques (e.g. average linkage method) or by selecting *k* items at random to serve as *seeds* for clusters (which are later replaced by centroids (mean vectors) of clusters).
2. If *seeds* are chosen, the remaining points in the data set are allocated to the cluster with the nearest *seed* (based on Euclidean distance); and as soon as a cluster has more than one member, the seed is replaced by its cluster centroid (Rencher, 2002). The mean or centroid of the *k* clusters is computed. The change in the clustering criteria (within-group sum of squares) produced is calculated when each object is moved from its cluster to another.
3. The change that leads to the greatest improvement in the clustering criteria (within-group sum of squares) is made.
4. Steps 2 and 3 are repeated until no more improvement in the clustering criteria (within-group sum of squares) occurs as a result of moving an object/individual/observation between clusters.

It is often difficult to decide on the number of groups and no method qualifies for recommendation in all circumstances (Everitt et al., 2001) and most methods are informal (Everitt & Dunn, 2001). The within-group sum of squares decreases as new clusters are added and ultimately takes a value of zero if every point is made into a cluster; therefore, a sharp change may be indicative of the best solution. One way of noting this reduction is to plot a graph between the number of clusters and Within-group sum of squares and look for an "elbow" in the curve (Everitt & Hothorn, 2006), which would be suggestive of a particular number of clusters (Everitt & Dunn, 2001). In K-Means clustering, when variables are on measured on different scales, some form of standardization of variables is necessary before applying the method (Everitt et al., 2001).

Before apply K-Means clustering, the variables *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* were standardized by dividing each variable by its respective range (Everitt & Hothorn, 2006). Figure 6.2 shows a graph between the number of clusters and Within-group sum of squares. As "little elbows" were noticed corresponding to 2 and 4 number of clusters, detailed K-means clustering was applied using 2 and 4 as the number of clusters to the data to ascertain groupings in each case. For the 2 cluster solution, Table 6.16 shows the number of drivers in each group and their group-mean values on the index *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings*. On average, group no. 1 has higher scores on the index *DPI3-weighted,* lower *No. of Total Hazards* and lower *No. of Low-speed Warnings* compared with group no. 2, which is a smaller group with 16 drivers. Drivers in group no. 2 may be considered as possessing poor driving skills. Although group no.2 is small, its size is still relatively large compared with the total size of the sample. Other ancillary data can be found in Appendix-E.

For the 4 cluster solution, Table 6.17 shows the number of drivers in each group and their group-mean values on the index *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings*. Inspection of table 6.17 indicates that there is no consistent pattern with regard to the driving performance skill of groups and their scores on the three variables e.g., group no.1 has the worst score on the *No. of Total Hazards,* whereas its score on the *No. of Low-speed Warnings* is not the worst. This inconsistent behavior was expected (as pointed out in Section 6.4), as subject knowledge indicated that more than 2 or 3 groups would not provide meaningful/relevant classifications.

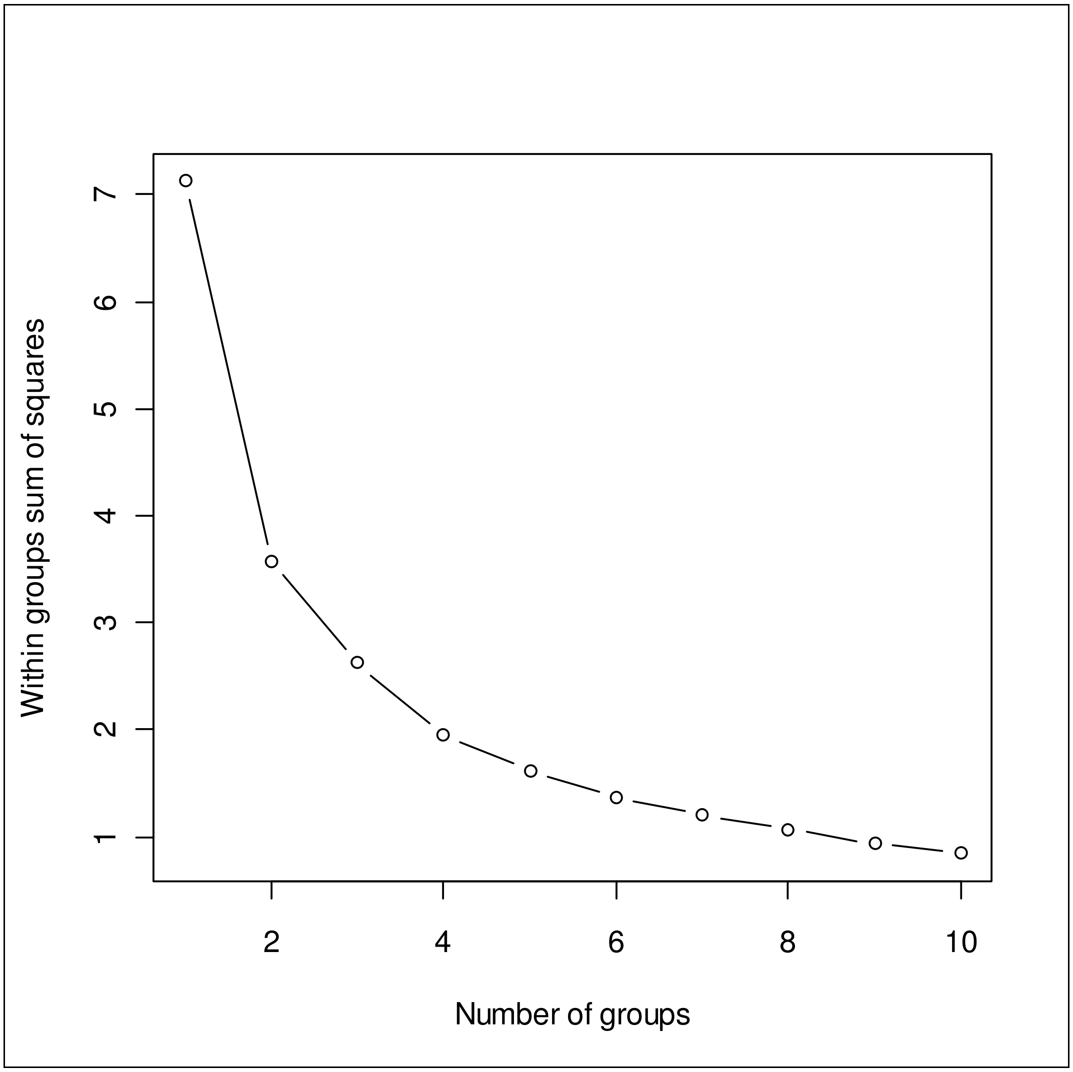



**Figure 6.2** Within groups sum of squares plotted against number of clusters using K-Means Clustering.

**Table 6.16** Mean values of *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* considering 2 groups in K-means Clustering.

| 2 Cluster/Group solution using K-means Clustering | | | | |
|---|---|---|---|---|
| Group No. | No. of Drivers | Mean *No. of Total Hazards* | Mean *No. of Low-Speed warnings* | Mean *DPI3-weighted* |
| 1 | 40 | 6.975 | 393.525 | 0.3595664 |
| 2 | 16 | 13.75 | 511.4375 | -0.898916 |

**Table 6.17** Mean values of *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* considering 4 groups in K-means Clustering.

| 4 Cluster/Group solution using K-means Clustering | | | | |
|---|---|---|---|---|
| Group No. | No. of Drivers | Mean *No. of Total Hazards* | Mean *No. of Low-Speed warnings* | Mean *DPI3-weighted* |
| 1 | 7 | 17.42857 | 429.7143 | -1.231837 |
| 2 | 8 | 11.25 | 588.125 | -0.6370448 |
| 3 | 23 | 8.043478 | 431.6087 | 0.1744398 |
| 4 | 18 | 5.666667 | 349.1111 | 0.5392833 |

## 6.5 Results

The driver groups possessing poor driving performance skills identified in Sections 6.4.1, 6.4.2, 6.4.3, 6.4.4, 6.4.5, 6.4.6 and 6.4.8 (two cluster case, k-means) were not relatively small compared with the total number of drivers, and hence the classifications obtained as a result of these scenarios were not viable as pointed in Section 6.4. The groupings obtained in Section 6.4.8 (four cluster case, k-means) were inconsistent with regard to the variables and therefore were not meaningful. The clustering results obtained using K-Means clustering (both 2 cluster case and 4 cluster case) were not promising. It may also be highlighted that scenarios in section 6.4.1, 6.4.2, 6.4.3 and 6.4.4 did not take into account the most intuitively appropriate 3 variables simultaneously (i.e. *DPI3-weighted* or *DPI3, No. of Total Hazards* and *No. of Low-speed Warnings*) that were relevant, keeping in view the psychometric principles used in the design of the drive in context of driving behavior exhibited by cognitively deficient drivers (see Sections 4.3.1, 4.4.2 and 5.4.2). The viability of the simultaneous use of these three variables is also supported by the fact that mixture-model based scenarios that used these variables simultaneously gave exactly the same driver groups and classification. The most intuitively appropriate and clinically relevant classification of drivers was provided by considering *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* with a 3 cluster solution (scenario of Section 6.4.7) using normal-mixture model EEV. It may be pointed out that age was not used as a factor in the decision making/model selection process. The following points support use of this model:

1. The poor-performance driver group (group no.2) was a relatively smaller group (8 drivers) compared with the total sample.
2. Despite the fact that drivers in Group no.2 were driving on average at the lowest speeds, they had the greatest number of accidents etc (i.e. no. of total hazards) and had low rating on all other driving performance measures (i.e. *DPI3-weighted*). This substantiates their classification as "poor drivers".
3. The total number of groups/clusters (i.e. three) was within the limits (2 or 3) of the total number of groups.
4. There was a consistent pattern with regard to the driving performance skill of groups and their scores on the three variables i.e., on average, group no.3 had the highest score on *DPI3-weighted*, the lowest *No. of Total Hazards* and the lowest *No. of Low-speed Warnings*. Similarly, group no.1 had favourable scores on all these variables compared with group no. 2.
5. The Maximum BIC occurred for the 2 cluster model (model VEI, BIC= -1070.690); the difference in BIC between the 2 cluster model (model VEI, BIC= -1070.690) and the three cluster model (model EEV, BIC= -1075.632) was small enough to conclude that there were either two or three groups in the data. The three group/cluster model (model EEV, BIC= -1075.632) provided the most clinically relevant classification. Adoption of this practice is as per recommendation of Fraley and Raftery (2002).
6. In order to evaluate the clusters obtained from our model, scores on other variables of interest (i.e. cognitive tests) were compared (Cohen et al., 1977) using driver classification groupings. Scores on the nine cognitive tests (*trail, clock, rey-copy, rey-recall, dichotic, paper, ufov1, ufov2, ufov3*) on average were the highest by group no.3, then group no. 1 and then group no. 2. This same order was also observed in decreasing order of driving-performance-skill among the three groups.
7. Five "ideal objects" (i.e., 2 older group drivers and 3 younger group drivers) were also part of the sample. These five drivers were "ideal" in the sense that their driving performance/skill was obvious (and well known) on account of their long association with the university/department. After cluster analysis, it was noted

that these five drivers ended up in the most appropriate groups and fitted in the classification very well. This procedure is as per recommendation of Gordon (Gordon, 1999).

It may be noted that despite the fact that drivers in Group no.2 were driving on average at the lowest speeds, they had the greatest number of accidents etc (i.e. no. of total hazards) and had low rating on all other driving performance measures (i.e. *DPI3-weighted*). This is contrary to a speed accuracy trade-off effect, which in a different context involving older drivers on a driving simulator was reported (Park et al., 2007; Quillian et al., 1999), in DAT subjects in road tests by Hunt et al. (1997a), in out-patients with Alzheimer's disease in a driving simulator by Cox et al. (1998), and in suspected dementia subjects in a driving simulator by Szlyk et al. (2002).

## 6.6 Summary

This chapter has described the process of categorization of drivers based on driving skill. The rationale for the use of the statistical technique of cluster analysis rather than a categorization that is based on the control group (and its associated 2 standard deviations method) has been explained. Since in cluster analysis, it is generally impossible a priori to perceive what combination of variables is likely to result in interesting/informative classifications, six different scenarios having different combination of driving performance indices and/or variables/parameters were considered for clustering. The different methods that are used to classify observations/objects into clusters/groups have been explained along with their pros and cons. The relatively recently developed normal-mixture model clustering technique has been described along with its favourable aspects compared with the other methods and especially its ability to discern between members of overlapping / intersecting clusters.

Cluster analysis considering six scenarios using normal-mixture models has been illustrated along with the use of K-means clustering on one particular scenario. The most appropriate, well-fitting and clinically relevant/meaningful clustering scheme was provided by the normal-mixture model EEV with 3 clusters by considering the index *DPI3-weighted* and the variables *No. of Total Hazards* and *No. of Low-speed Warnings*.

This clustering scheme identified 8 drivers as possessing poor driving skills that belonged to the older driver group. Despite the fact that these drivers were on average driving at the lowest speeds, they had the greatest number of accidents etc (i.e. no. of total hazards) and had lowest rating on all other driving performance measures (i.e. *DPI3-weighted*), on average. The next chapter describes multiple regression analysis in modeling driving performance using cognitive tests (independent variables).

# 7 Development of Linear Models

## 7.1 Introduction

The objectives of this chapter were to develop parsimonious multiple linear regression models using nine neuropsychological tests (t*rail, clock, rey-copy, rey-recall, dichotic, paper, ufov1, ufov2, ufov3*) as predictors in order to predict general driving ability. In Chapter 5, six different driving performance indices were developed through the concept of Scale Development. These were *DPI1*, *DPI2* and *DPI3* and their corresponding weighted versions *DPI1-weighted*, *DPI2-weighted*, and *DPI3-weighted*. To model (linear regression model) the driving performance indices through the different cognitive tests it was necessary that the indices included the effects of all viable driving performance parameters i.e. the indices that were derived by considering all 24 driving performance parameters, which were *DPI1* and *DPI1-weighted*. Since the Pearson Moment Correlation coefficient between *DPI1* and *DPI1-weighted* was very high (0.9975), it was decided to model *DPI1-weighted* (dependent variable) through cognitive/neuropsychological tests (independent variables), as this index was geared to provide maximum discrimination between drivers because of Principal Component weights. The following sections describe the details of the development of linear regression models.

## 7.2 Screening of Variables

There were nine neuropsychological tests (*trail, clock, rey-copy, rey-recall, dichotic, paper, ufov1, ufov2, ufov3*) i.e., nine independent variables and 56 observations. According to Kutner et al. (2005), as a general rule of thumb, there should be at least 6 to 10 cases (observations) per variable, whereas, Harrell et al. (1996) recommend that if the number of predictors is > n/10 (where n is the number of observations/cases), a data reduction technique should be used to reduce the number of candidate predictors to enhance the accuracy/reliability of the model. For small ratios of the number of observations to the number of predictors, it is risky to generalize regression results beyond the sample because the regression coefficients are unstable due to their large standard errors. Therefore, screening of candidate predictors was necessary for early removal of predictors that had little chance of being predictive. In screening predictors, the outcome variable (i.e., *DPI1-weighted*) was not utilized (Harrell et al., 1996) but rather subject matter knowledge was used as a guide in screening predictors. The UFOV test consisted of three subtests (*ufov1, ufov2* and *ufov3*). In the first subtest the examinee is to identify a target (silhouette of a car or truck) that is presented in a central fixation box for varying lengths of time. In the second subtest, the examinee has to identify a central target (silhouette of a car or truck presented in the central fixation box) and also to radially localize a simultaneously presented target (silhouette of a car) displayed in any one of eight radial positions on the periphery of the computer monitor. The third subtest is identical to the second subtest except that the peripheral target (Silhouette of a car) is embedded in distractors (47 triangles of the same size and luminance as the peripheral target). Therefore, all features of subtest 1 (i.e. *ufov1*) are present in subtest2 (i.e. *ufov2*) and all features of subtest2 are present in subtest3 (i.e. *ufov3*). Hence, *ufov3* has all the features of *ufov1* and *ufov2*. This is also evident from the correlations coefficients (Table 7.1) between *ufov1, ufov2* and *ufov3* which progressively increase between these three subtests as there is more in “common” between *ufov2* and *ufov3* than between *ufov1* and *ufov2*. Also, the difficulty of the subtests progressively increases from *ufov1* to *ufov3.* Therefore, in view of the above the predictors *ufov1* and *ufov2* were dropped.

**Table 7.1** Correlation coefficients between the three subtests of UFOV test

| | *ufov1* | *ufov2* | *ufov3* |
|---|---|---|---|
| *ufov1* | 1.0000 | | |
| *ufov2* | 0.4225 | 1.000 | |
| *ufov3* | 0.3204 | 0.7471 | 1.000 |

The scores on the clock drawing test could range from 0 to 15. Out of the 56 clock drawing scores, 1 driver achieved a score of 10, 9 drivers achieved a score of 14 and 46 drivers achieved a score of 15. Due to the lack of variability this test was not discriminatory, as a variable which cannot vary will not co-vary (with the dependent variable) as well. Since the *clock* scores varied over a small range, they were bound to show no effect and were deemed non-significant (Draper & Smith, 1998; Montgomery et al., 2006). Therefore, it was also decided to drop the predictor *clock*, thus, finally resulting in the following six candidate predictors: *trail, rey-copy, rey-recall, dichotic, paper* and *ufov3*. Graphs of the nine cognitive tests against age with superimposed Lowess curves are are shown in Figure 7.1, 7.2 and 7.3. The lack of variability of the Clock-drawing test is obvious from Figure 7.1, where most of the participants had a score of 14 or 15 with one individual having a score of 10. In general, the cognitive tests show decrements in performance with respect to age.

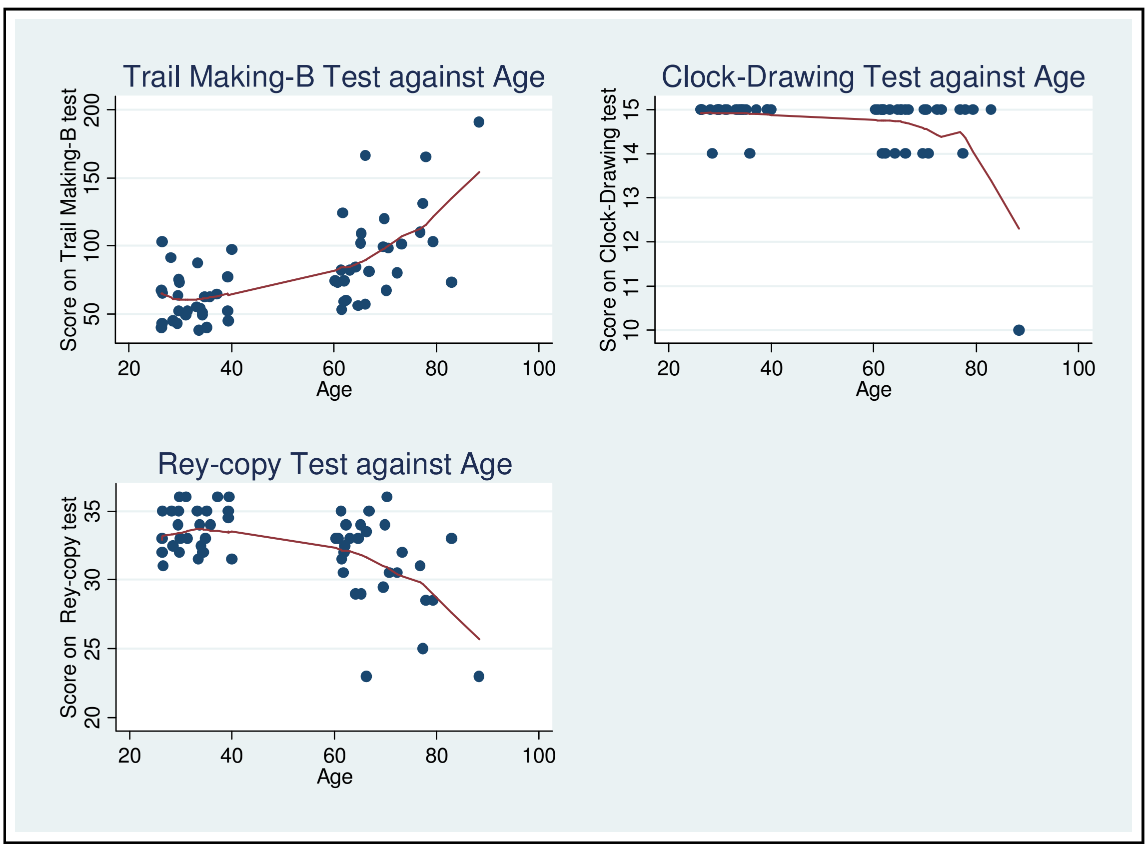



**Figure 7.1** Graphs of scores on the Trail Making-B Test, the Clock-drawing Test and the Rey-copy Test against Age, with superimposed Lowess curves.

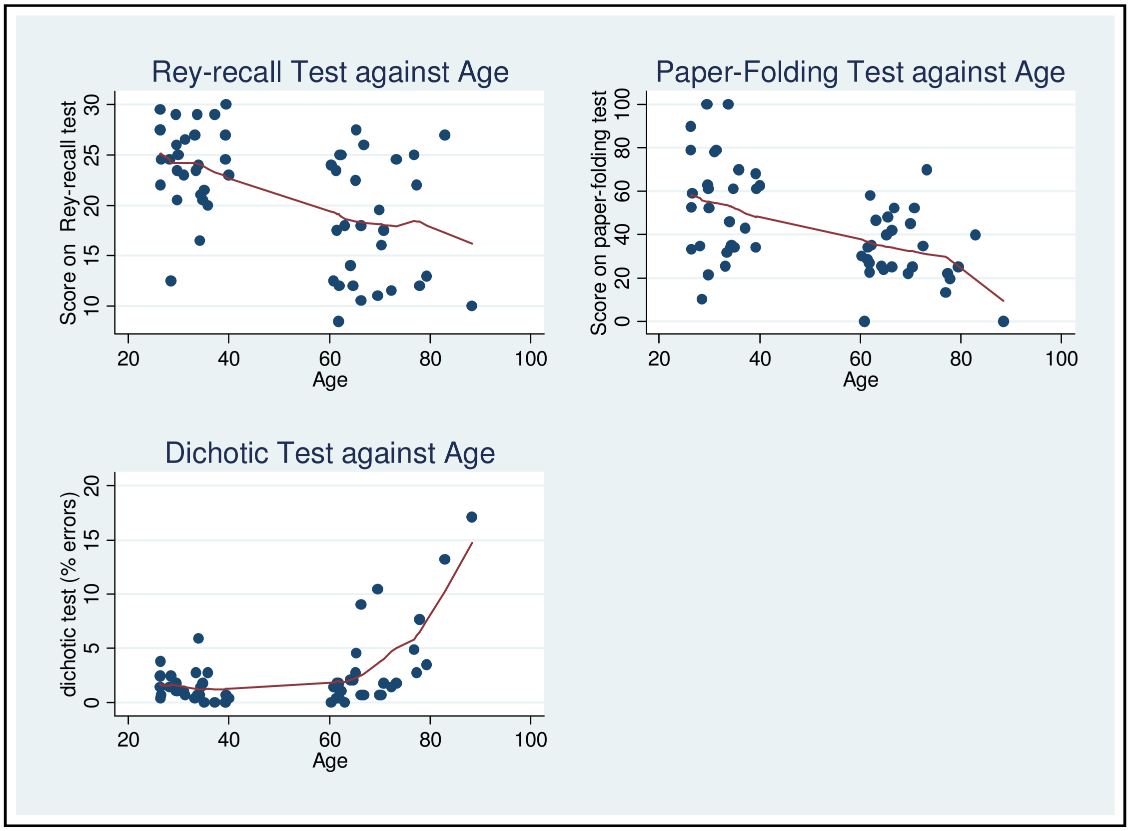



**Figure 7.2** Graphs of scores on the Rey-recall Test, the Paper-Folding Test and the Dichotic Listening Test against Age, with superimposed Lowess curves.

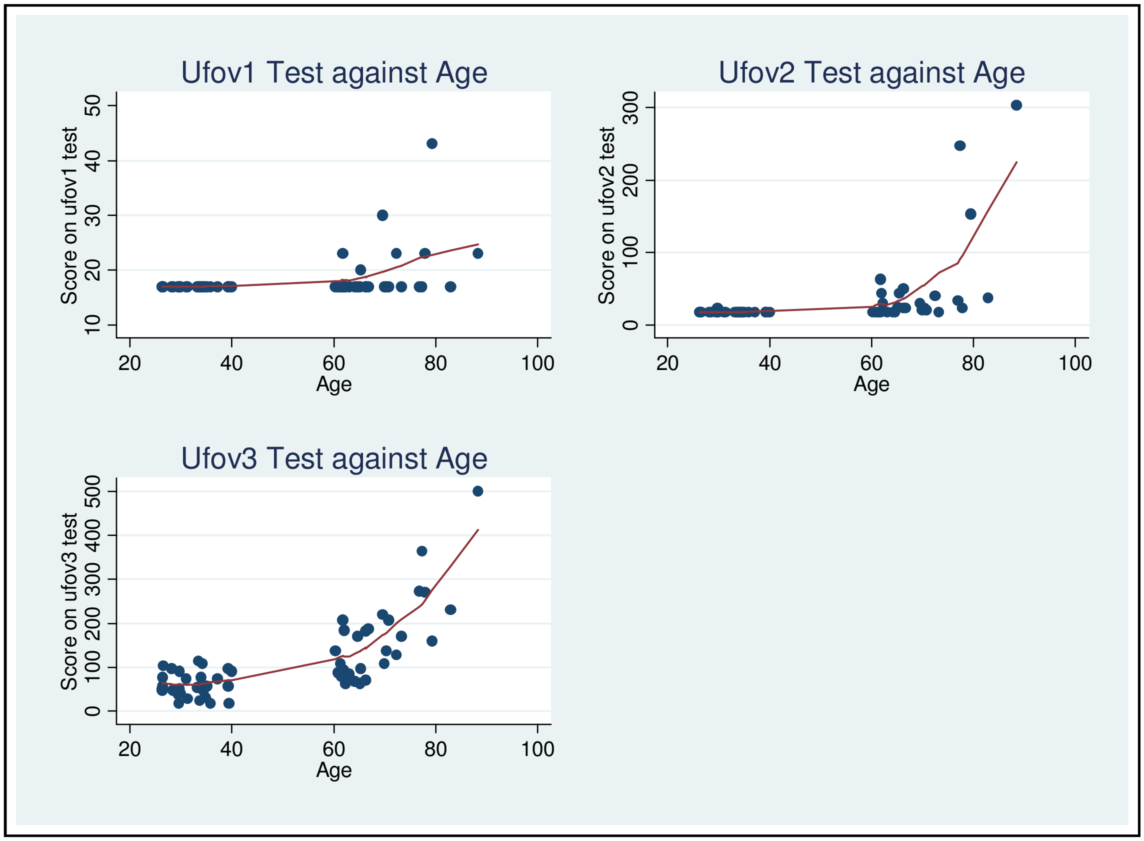



**Figure 7.3** Graphs of scores on the three subtests of the UFOV Test against Age, with superimposed Lowess curves.

## 7.3 Missing Value Problem

Driver bearing identification code "O56" (age 88.4 yrs) attempted the dichotic listening test but due to his inability to discern between the signals of the left and right channels (ears), ended up with a missing value score on this test. Since the case was very crucial because he was the oldest driver and had the highest *No. of Total Hazards* (i.e., thirty) and was the only individual with relatively substantially low score on the clock drawing test (score of 10),which is frequently used in dementia evaluations, it was decided to employ an appropriate methodology for addressing the missing value problem. In such cases, Altman (1991) suggests the following approach:

- Assign the most optimistic outcome to the missing value and analyze the data.

- Repeat the analysis using the most pessimistic outcome for the missing value.
- Reanalyze the data by dropping the missing value case.

The most pessimistic value was arrived at by assigning a score of 100 percent total error to driver "O56" on the dichotic listening test, since the driver had exhibited complete inability to take the test. The most optimistic value was arrived at by considering the estimation of the missing value by one of the missing-value-estimation-methods. The ad hoc method of mean substitution, where the mean of the variable (*dichotic*) is substituted for the missing value was not favoured as it artificially reduces the variance of the relevant variable and also diminishes relationships with other variables i.e. it distorts the covariance structure (Wayman, 2003). Another ad hoc method, regression predictions tends to artificially inflate correlations (Schafer, 1997) and was therefore ruled out. The procedure outlined by Schafer (1997) and Schafer & Olsen (1998) that utilizes the Expectation and Maximization algorithm (Dempster et al., 1977 cited in Schafer & Olsen, 1998) and multiple imputation (Rubin, 1987 cited in Schafer & Olsen, 1998) was employed. The procedure is based on the multivariate normal distribution model. Details of the method can be found in Schafer (1997) and Schafer & Olsen (1998). A free downloadable software program by the name of NORM which is Windows® based is available at the Pennsylvania State University website at http://www.stat.psu.edu/~jls/misoftwa.html. In multiple imputation (MI), each missing value is replaced by a set of m > 1 plausible values, which are drawn from their predictive distribution; typically 3 to 5 imputations (i.e., m is from 3 to 5) are sufficient for good results. After Multiple Imputation, the m apparently complete data sets are analyzed by complete-data methods. After each of the m data sets is subjected to identical analysis, the results (i.e., estimates and standard errors) are combined using rules provided by Rubin (Rubin, 1987 cited in Schafer & Olsen, 1998) and others, to finally present overall estimates and standard errors that reflect missing-data uncertainty. However, it was not worth the computational effort/precision in our case to carry out m imputations as assigning the most optimistic outcome to the missing value based on missing-value-estimation-methods was indeed rather very optimistic as driver "O56" had failed to take the test due to cognitive deficits / lack of divided attention abilities. Therefore, instead of m imputations, the NORM software was used to make only one imputation (Tabachnick & Fidell, 2007), which tended to provide a better estimate of the missing value than the ad hoc methods (Little & Rubin, 1987; Graham et al., 1994 cited in Schafer & Olsen, 1998).

Variables possessing heavily skewed distributions were first transformed to approximate normality and then transformed back to their original scale after the imputation. Table 7.2 shows the different transformations used to bring about approximate normality. Since the lowest value of *dichotic* was zero, a small increment (i.e. 1) was added to the *dichotic* test result before applying the transformation as it is necessary that a variable should not be negative or zero before the power transformation is applied.

**Table 7.2** Different transformations used before imputation

| Variable | Transformation |
|---|---|
| *trail* | $\frac{1}{\sqrt{trail}}$ |
| *rey-copy* | $(rey-copy)^3$ |
| *rey-recall* | $(rey-recall)^3$ |
| *dichotic* | $\frac{1}{\sqrt{(dichotic+1)}}$ |
| *paper* | none |
| *ufov3* | $\log_e(ufov3)$ |

The particular transformations were arrived at by using the *ladder of power* function in the statistical software Stata® and by inspection of histograms after transformation of the variables. This function searches for a subset of powers for a transformation that converts a variable into a normally distributed variable by comparing it with the quantiles of a normal distribution. The Expectation Maximization (EM) algorithm is a general technique that is used to fit models in incomplete data. It uses the relationship between missing data and the unknown parameters of a model. Using an iterative approach, first values of the parameters of the model are assumed and the missing values are predicted, then these predictions are used to update the parameter estimates and so on. This process ultimately converges to maximum-likelihood estimates that implicitly average over the distribution of the missing values (Schafer & Olsen, 1998). Using the NORM software the Expectation and Maximization (EM) algorithm converged after 8 iterations. The number of iterations input was 10000 (using a seed no. 99086 for reproducibility) and the software was instructed to impute at every $10000^{th}$ iteration, which resulted in a single imputed value of 17.11. Therefore, the most optimistic value assigned to the missing value for the *dichotic* variable was a total error rate of 17.11 percent on the dichotic listening test.

## 7.4 Scenarios for Regression modelling

Owing to the missing value encountered for the dichotic listening test with driver "O56", linear regression models were developed for the following three scenarios:

1. Pessimistic Case (dichotic score = 100 for driver "O56").
2. Optimistic Case (dichotic score = 17.11 for driver "O56").
3. Deletion of Case of Driver "O56".

The dependent variable was *DPI1-weighted* and the independent variables (predictors) were the cognitive/neuropsychological tests: *trail, rey-copy, rey-recall, dichotic, paper* and *ufov3*. First full models incorporating all six predictors for all three scenarios were developed and its diagnostics performed. Then the computationally intensive technique of *all regression models* was used on each scenario for the selection of the best model. Finally, these best models were subjected to diagnostic checks and confirmed.

### 7.4.1 Full Model Development

Details of multiple linear regression modeling can be found in Weisberg (2005), Draper & Smith (1998), Montgomery et al. (2006), Kutner et al. (2005), Cohen et al. (2003), Berk (2004), Christensen (2001) and Faraway(2005). Full models employing all six cognitive tests as predictor were first developed for each of the three scenarios of Section 7.4. The first step was to explore non linear transformations on the predictors and/or the dependent variable. The main purpose of the non linear transformation is to make the mean function linear in the transformed scale. The scatter matrix plot between the six predictors and *DPI1-weighted* showed no evidence of straight line mean function, therefore non linear transformation of the predictors and/or dependent variable was deemed necessary. With one predictor and one response variable, a suitable transformation can be selected by visualizing the mean function in a scatter plot and ensuring that the resulting scatter plot has an approximate straight line mean function. With many predictors, this visualization technique is not effective. According to Weisberg (2005), the overall objective of transformation is that the multiple linear regression matches the data to a good approximation. Also, in a large number of cases, linearity and variance stabilization can be achieved by employing the same transformation (Cook & Weisberg, 1999). When there is more than one predictor, all transformations for the predictors are chosen at the same time so as to make the *joint* distribution of the predictors as normal (i.e. multinormality) as possible and then the Box and Cox procedure (Box & Cox 1964 cited in Weisberg, 2005) is used on the transformed predictors to find a transformation for the response variable (Cook & Weisberg, 1999; Weisberg, 2005). To find the transformations that do the best job at achieving joint normality of the predictors, the method of maximum likelihood is used. The *power family* of transformations are most often used in this context, which is given by:

$$\psi(U,\lambda) = U^{\lambda} \tag{7.1}$$

Where $U$ is the variable of interest and $\lambda$ is the power parameter which is varied in order to get members of this family. For example, the square root and the cube root transformation corresponds to $\lambda = \frac{1}{2}$ and $\lambda = \frac{1}{3}$ respectively. $\lambda = 0$ is interpreted as a log transformation and $\lambda = 1$ corresponds to no transformation. Values of $\lambda$ that are considered, vary from -2 to 2 (Draper & Smith, 1998). For these transformations to be used, it is necessary that the variable $U$ be strictly positive; this can be achieved by adding a small increment such that the minimum value of the variable becomes positive. It essentially slightly translates the scale without changing its distribution. If the ratio of the maximum to the minimum value of a variable is less than 10, then any transformation of the variable is unlikely to be helpful (Draper & Smith, 1998; Weisberg, 2005; Cohen et al., 2003), because the relationship between the variable and its transformed from will essentially be linear. Since the ratio of the maximum to the minimum value of the predictors *trail, rey-copy* and *rey-recall* was less than 10 for all three scenarios, it was decided not to consider them for transformation. The predictors *dichotic* and *paper* had a minimum value of zero (i.e. non-positive) for all three scenarios, therefore a small increment i.e. 1 was added to them to make them positive. The response variable *DPI1-weighted* was non-positive as it ranged from -2.129 to 1.154 for the *Pessimistic Case* and *Optimistic Case* and from-1.368 to 1.154 for the *Deletion of Case of Driver "O56"* scenario, the following formula (Streiner and Norman, 2003) was used to convert it into positive values called T-score given by the equation:

$$\text{Tscore\_DPI1\_weighted} = \bar{X}' + z(SD') \tag{7.2}$$

Where,

$\bar{X}'$ = desired mean of the variable i.e.=100

$SD'$ = desired standard deviation of the variable i.e.=15

$z$ = the original standardized score of *DPI1-weighted* (i.e. subtracting the mean from each *DPI1-weighted* value and then dividing by its standard deviation, so that mean of *DPI1-weighted* is zero and its standard deviation is 1).

Tscore_DPI1_weighted = *DPI1-weighted* converted to T-score such that its mean is now 100 and its standard deviation is 15.

A T-score is simply a z-score (i.e. standardized score) with a new mean (i.e. 100) and standard deviation (i.e. 15). The choice of the mean of 100 and standard deviation of 15 was to make sure that we did not end up with negative values on Tscore_DPI1_weighted, as before considering variables as candidates for power transformations, it is necessary that they be positive (Weisberg, 2005). Equation 7.2 is a linear transformation and does not affect the distribution of *DPI1-weighted*. The three predictors *dichotic+1*, *paper+1* and *ufov3* were considered candidates for transformation in all three scenarios of Section 7.4. The Statistical software R® that is based on the S programming language with the add-on package *alr3* containing the function *bctrans* was used to find the powers $\lambda$ to bring about joint normality using the method of maximum likelihood.

**Table 7.3** Summary table showing estimated $\lambda$ and value adopted for use for the three scenarios.

| Variable | Pessimistic Case | | Optimistic Case | | Deletion of Case of Driver "O56" | |
|---|---|---|---|---|---|---|
| | Estimated $\lambda$ | $\lambda$ rounded for use | Estimated $\lambda$ | $\lambda$ rounded for use | Estimated $\lambda$ | $\lambda$ rounded for use |
| *(dichotic + 1)* | -0.4851 | -0.5 | -0.3895 | -0.5 | -0.3965 | -0.5 |
| *(paper + 1)* | 0.6777 | 0.5 | 0.6830 | 0.5 | 0.6875 | 0.5 |
| *ufov3* | 0.1640 | 0 | 0.1544 | 0 | 0.1797 | 0 |

Table7.3 shows the detail of estimated values of $\lambda$ for each scenario and the value adopted after rounding. Rounding is done because as practical matter, very precise estimates of $\lambda$ are not needed and the rounded figure will be more likely to be meaningful in any real-life situation (Weisberg, 2005; Cook & Weisberg, 1999). After rounding, the adequacy of the rounded values of $\lambda$ was tested through likelihood ratio tests which gave p-values of 0.33, 0.29 and 0.36 for the three scenarios. $\lambda = 0$ is interpreted as a log transformation. Next the six predictors (including three transformed) i.e., $(dichotic+1)^{-0.5}$ , $(paper+1)^{0.5}$, $\log_e ufov3$, $trail$, $rey-copy$ and $rey-recall$ were used in a Box and Cox procedure (Box & Cox 1964 cited in Weisberg, 2005) to find transformation of the response variable Tscore_DPI1_weighted using the statistical software Stata®. Table 7.4 gives values for $\lambda$ along with 95% confidence intervals for the three scenarios for the response variable Tscore_DPI1_weighted .

**Table7.4** Summary table showing estimated $\lambda$ along with 95% confidence intervals for $\lambda$ for the three scenarios for the response variable $\mathrm{Tscore\_DPI1\_weighted}$ .

| Pessimistic Case | | Optimistic Case | | Deletion of Case of Driver "O56" | |
|---|---|---|---|---|---|
| Estimated $\lambda$ | 95% confidence interval | Estimated $\lambda$ | 95% confidence interval | Estimated $\lambda$ | 95% confidence interval |
| 2.218 | 0.75 to 3.67 | 2.289 | 0.83 to 3.74 | 2.557 | 1.06 to 4.05 |

Since the confidence intervals for all three scenarios were wide that included two or more of the bench mark levels of $\lambda = -1, -\frac{1}{2}, 0, \frac{1}{2}, 1$ which indicated that $\lambda$ was not crisply estimated. It implied that it made little difference which of the wide range of values was used and since the wide confidence intervals included 1, it would not be worthwhile to transform the response variable (Draper & Smith, 1998). According to Cook and Weisberg (1999), estimates of $\lambda$ outside the range of -2 and +2 often indicate that transformation of the response will not be helpful. Therefore, it was decided to use no transformation for the response variable $\mathrm{Tscore\_DPI1\_weighted}$ for all three scenarios. The following six predictors were regressed against $\mathrm{Tscore\_DPI1\_weighted}$ :

Trail = $trail$

Rey_copy = $rey - copy$

Rey_recall = $rey - recall$

T_dichotic_t_plus = $\left(dichotic + 1\right)^{-0.5}$

T_paper_plus = $\left(paper + 1\right)^{0.5}$

Ln_ufov3 = $\log_e ufov3$

Results of regression analysis for the *Pessimistic Case* are shown in Table 7.5 to Table 7.7.

**Table 7.5** Detail of Sum of Squares for the scenario *Pessimistic Case.*

| Source | Sum of Squares | Degrees of freedom | MS |
|---|---|---|---|
| Model | 7721.66822 | 6 | 1286.9447 |
| Residual | 4653.33179 | 49 | 94.96 |
| Total | 12375 | 55 | 225 |

**Table 7.6** Detail of F-test and R-squared for the scenario *Pessimistic Case.*

| Detail of F-Test and R-Square | |
|---|---|
| Number of observations | 56 |
| F (6, 49) | 13.55 |
| Prob>F | 0.0000 |
| R-squared | 0.6240 |
| Adj R-squared | 0.5779 |
| Root MSE | 9.745 |

**Table 7.7** Detail of estimated coefficients and their standard errors for the scenario *Pessimistic Case.*

| Predictors | Coefficient | Standard Error | t | P>\| t \| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| Trail | -0.051291 | 0.069930 | -0.73 | 0.467 | -0.1918 | 0.08924 |
| rey_copy | 0.438801 | 0.821197 | 0.53 | 0.596 | -1.2114 | 2.08906 |
| rey_recall | 0.402275 | 0.308616 | 1.30 | 0.199 | -0.2179 | 1.02246 |
| T_dichotic_t_plus | 17.09213 | 8.676566 | 1.97 | 0.055 | -0.3440 | 34.5283 |
| T_paper_plus | 1.538772 | 1.017106 | 1.51 | 0.137 | -0.5051 | 3.58272 |
| Ln_ufov3 | -5.026747 | 2.544358 | -1.98 | 0.054 | -10.139 | 0.08633 |
| constant | 82.63642 | 30.96989 | 2.67 | 0.010 | 20.4001 | 144.8727 |

The test for significance of regression is significant with $F(6, 49) = 13.55$, $p < 0.00005$ showing that at least one of the predictor variable contributes significantly to the model. Results of regression analysis for the *Optimistic Case* and for the *Deletion of Case of Driver "O56"* are in Appendix-F. For the *Optimistic Case*, the test for significance of regression is significant with $F(6, 49) = 13.42$, $p < 0.00005$ showing that at least one of the predictor variable contributes significantly to the model. For the *Deletion of Case of Driver "O56"* scenario, the test for significance of regression is significant with $F(6, 48) = 9.66$, $p < 0.00005$ showing that at least one of the predictor variable contributes significantly to the model.

### 7.4.2 Full Model Diagnostics

After full multiple linear regression models were developed for the three scenarios of Section 7.4, the next step was to subject them to a thorough analysis to ensure that the functional form of the models was correct, the assumptions were satisfied and to identify outliers/influential observations. The diagnostics were primarily related to the analysis of residuals so as to determine the adequacy of the model. Once the adequacy of the model had been ensured, the next step was to select parsimonious models. The following paragraphs describe the diagnostics for the *Pessimistic Case* (unless otherwise produced in this section, all graphs and tables are in Appendix-F). First, the residuals were plotted against the predicted value (yhat) and a Lowess smoother was superimposed on the graph. It may be highlighted that the Lowess curve follows the general trend in the data. The residuals were distributed symmetrically around the zero-residual line with no evidence of curvilinearity. The Lowess curve did not exhibit any large or systematic deviations (or curvature) from the zero-residual line showing that the functional form of the model was correct. It may be noted that the Lowess curve's downturn at the ends should be disregarded as a Lowess artifact, because there are only a few cases in the extremes that determine its position/location; rather we should concentrate more on the central parts of the curve (Weisberg, 2005; Hamilton, 2005).

Apart from the visual inspection of residuals for ensuring equal variance, two separate statistical tests for heteroskedasticity of residuals (i.e. residuals have different variances) were performed using the statistical software Stata®: (1) Cameron & Trevedi's decomposition of IM-test, and (2) Breusch-Pagan/Cook-Weisberg test of heteroskedasticity. Both tests test the hypothesis: $H_o$= Residuals have constant/equal variance (i.e. homoskedasticity of residuals is present). P-vales of 0.2276 and 0.6811 were obtained as evidence of homoskedasticity of residuals (i.e. heteroskedasticity in residuals was absent), and therefore this important assumption was satisfied. Plots of residuals against each of the six predictors with superimposed Lowess curve can also be helpful in indicating model inadequacies; such plots did not show any systematic pattern i.e., the Lowess fitted line did not show any large or systematic deviations from the zero-residual line, thus manifesting the adequacy of the functional form of the model. Normality of residuals was checked through two plots: (1) kernel density plot (with superimposed normal curve), and (2) Q-Q plot. The kernel density plot is a better alternative to the histogram and approximates the probability density of a variable (residuals). The Q-Q plot is a plot of the quantiles of the residuals against the quantiles of the normal distribution and is sensitive to non-normality near the tails. The kernel density plot matched the normal curve quite satisfactorily and so did all the points (except one point) lie on a straight line in the Q-Q plot. Besides, the residuals were subjected to three statistical tests of normality in Stata® (Shapiro-Wilks test, Shapiro-Francia test and Skewness/Kurtosis test), which gave p-values of 0.19449, 0.12764 and 0.0636 respectively ($H_o$= residuals are normally distributed) as evidence of normality of residuals.

Augmented component plus residual plot of each of the six predictors was also plotted with Lowess and fitted lines. These plots are used to diagnose nonlinearities and suggest alternate functional form (Hamilton, 2005). The Lowess curves did not show any systematic pattern and closely followed the regression model thus reinforcing the conclusion that was reached from the residual versus predicted value (yhat) plot that the regression model adequately accounted for all nonlinearity in the data. It may be reiterated that the Lowess curve's downturn at the ends should be disregarded as a Lowess artifact, as they are less reliable at the edges of the plot.

Outliers are identified with respect to their y-values and/or their x-values. Outliers with respect to their y-values will have large studentized residuals and with respect to their x-values will have high leverages. A large value of $h_{ii}$ (diagonal element of the hat matrix i.e. leverage) indicates that the $i^{th}$ case is located further away from the center of all x observations. A large leverage can result from extreme x values or unusual combination of x values (i.e. although none of the individual x values is unusual by itself). Not all outlying cases are influential. An observation is influential if its exclusion causes major changes in the fitted model (estimated coefficients, fitted values, t-tests, etc). The three measures of influence that are widely used in practice are DFFITS, Cook's distance and DFBETAS (Kutner et al.,2005) and all measure influence based on omission of a single case. DFFITS and Cook's distance are closely related and tend to flag the same observations as influential (Hamilton, 2005; Cohen et al., 2003). The numeric values of the studentized residuals are less important than their pattern (Chatterjee & Hadi, 2006). A case having a large studentized residual may be a perfectly plausible observation and deleting it to improve the fit of the model is dangerous as it gives a false sense of precision in estimation or prediction (Montgomery et al., 2006). According to Chatterjee (Chatterjee & Hadi, 2006), cases with high leverages that are not influential are not a source of problems but high leverages cases that are influential should be scrutinized as these cases are outlying in the x-space (i.e., predictor space) and also influence the fit of the model. Added-variable plots (see Appendix-F) have slopes that are equal to the respective partial regression coefficients. If the slope of the best fitting regression line in the Added-variable plot is zero, the predictor variable has no unique relationship to the dependent variable; if the slope is positive, the added variable will have a positive relationship to the dependent variable and if it is negative, it will have a negative relationship (Cohen et al., 2003). Added-variable plots also help visually unveil observations that exert a disproportionate influence on the model (Hamilton, 2005). Points/observations that have high leverage are horizontally distant/distant from the rest of the data in Added-variable plots (Hamilton, 2005).

DFFITS measures the influence that case i has on the fitted value $\hat{Y}_i$ and is given by (Kutner et al.,2005):

$$(DFFITS)_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{MSE_{(i)} h_{ii}}} \tag{7.3}$$

Where,

$\hat{Y}_i$ = Fitted value for the $i^{th}$ case when all n cases are used in fitting the model.

$\hat{Y}_{i(i)}$ = Predicted value for the i^th^ case obtained when the i^th^ case is omitted in fitting the model.

The denominator in Equation 7.3 is used for standardization, so that the value $(DFFITS)_i$ for the $\text{i}^{\text{th}}$ case represents the number of estimated standard deviations of $\hat{Y}_i$ that the fitted value $\hat{Y}_i$ either increases or decreases by omitting the $\text{i}^{\text{th}}$ case in the regression model. As a guideline in identifying influential cases, if the absolute value of DFFITS exceeds 1 for small to medium data sets (Kutner et al.,2005) the case may be considered as influential.

DFBETAS is a measure of influence that case i exerts on the regression coefficients $b_k$ ($k$ = 0, 1, 2…, $p$-1) and is given by (Kutner et al.,2005):

$$(DFBETAS)_{k(i)} = \frac{b_k - b_{k(i)}}{\sqrt{MSE_{(i)} c_{kk}}} \tag{7.4}$$

Where,

$b_k$ = Regression coefficient based on all n cases.

$b_{k(i)}$ = Regression coefficient obtained when the $\text{i}^{\text{th}}$ case is omitted.

The denominator in Equation 7.4 is used for standardization, so that the value $(DFBETAS)_{k(i)}$ for the $\text{i}^{\text{th}}$ case represents the number of estimated standard deviations of $b_k$ that the coefficient $b_k$ either increases or decreases by omitting the $\text{i}^{\text{th}}$ case in the regression model. As a guideline in identifying influential cases, if the absolute value of DFBETAS exceeds 1 for small to medium data sets (Kutner et al.,2005) the case may be considered as influential.

The leverage versus squared residual plot is used to identify individual outliers. The square of the residuals enhances the effect of the residuals visually. A horizontal line depicting the mean of leverages and a vertical line depicting the mean of squared residuals are embedded for reference. Based on its particular combination of x values, the leverage shows how much potential an observation has for influencing a regression and large squared residuals point out observations whose observed y values are much different from that predicted by the model. Leverages greater than $\frac{2k}{n}$ (where, $k$ is the no.

of parameters in the model (including constant) and $n$ is the number of observations) are considered high (Kutner et al.,2005; Hamilton, 2005). In the *Pessimistic Case*, from the leverages versus squared residuals plot, drivers “O6”, “O10”, “O34”, “O39” and “O56” had high leverages (greater than $\frac{2k}{n} = 0.25$) and driver “O46” had a high studentized residual (-3.85) showing an ill-fit. Absolute values of Studentized residuals greater than 2 or 3 are considered high by Chatterjee and Hadi (Chatterjee & Hadi, 2006), while Montgomery et al. (2006) consider 3 or 4 to be the threshold. Based on the six cognitive tests, the value predicted by the model (i.e., y hat) for Driver “O46” was higher (and hence the negative studentized residual) than her observed value on the index, because her scores on the six cognitive tests were relatively favourable compared to the other older drivers while in actual practice her performance measured via the index was relatively poor. All drivers possessing high leverages were older drivers. Only one driver “O46” had an absolute DFFITS value (-1.43) greater than 1 and there was no absolute value of DFBETAS greater than 1, thus manifesting that driver “O46” was influential. This is also obvious from the Added-variable plot as driver “O46” is quite distant from the rest. However, as pointed out by Chatterjee & Hadi (2006) cases with high leverages that are not influential are not a source of problems but high leverages cases that are influential should be scrutinized as these cases are outlying in the x-space (i.e., predictor space) and also influence the fit of the model; since the high leverage cases were not influential therefore, there was nothing wrong with observation of driver “O46” and was left intact as it did not affect the coefficient estimates (i.e. all absolute values of DFBETAS less than 1). The outliers i.e. observations having high leverages and high residuals and also the influential cases were scrutinized to ensure that there was no measurement, recording or calculation error. After thorough checking, no error of any sort could be detected manifesting that these were indeed plausible observations. Therefore, the model for the *Pessimistic Case* was considered to be representative of all of the observations in the sample and not an artifact of a few.

For the *Optimistic Case* scenario (Appendix-F), all assumptions were satisfied through plots and statistical tests as in the *Pessimistic Case* scenario. In the *Optimistic Case* scenario, from the leverages versus squared residuals plot, drivers “O6”, “O34”, “O39”, and “O56” had high leverages (greater than $\frac{2k}{n} = 0.25$) and driver “O46” had a high studentized residual (-3.84) showing an ill-fit. Based on the six cognitive tests, the value predicted by the model (i.e., y hat) for Driver “O46” was higher (and hence the negative studentized residual) than her observed value on the index, because her scores on the six cognitive tests were relatively favourable compared to the other older drivers while in actual practice her performance measured via the index was relatively poor. All drivers possessing high leverages were older drivers. Only one driver “O46” had an absolute DFFITS value (-1.43) greater than 1 and there was no absolute value of DFBETAS greater than 1, thus manifesting that driver “O46” was influential. This is also obvious from the Added-variable plot as driver “O46” is quite distant from the rest. Since the high leverage cases were not influential therefore, there was nothing wrong with observation of driver “O46” and was left intact as it did not affect the coefficient estimates (i.e. all absolute values of DFBETAS less than 1) (Chatterjee & Hadi, 2006). The outliers i.e. observations having high leverages and high residuals and also the influential cases were scrutinized to ensure that there was no measurement, recording or calculation error. After thorough checking, no error of any sort could be detected manifesting that these were indeed plausible observations. Therefore, the model for the *Optimistic Case* was considered to be representative of all of the observations in the sample and not an artifact of a few.

For the *Deletion of Case of Driver “O56*” scenario (Appendix-F), all assumptions were satisfied through plots and statistical tests as in the *Pessimistic Case* scenario. In the *Deletion of Case of Driver “O56*” scenario, from the leverages versus squared residuals plot, drivers “O6”, “O10”, “O34”, and “O39” had high leverages (greater than $\frac{2k}{n} = 0.25$) and driver “O46” had a high studentized residual (-3.79) showing an ill-fit. Based on the six cognitive tests, the value predicted by the model (i.e., y hat) for Driver “O46” was higher (and hence the negative studentized residual) than her observed value on the index, because her scores on the six cognitive tests were relatively favourable compared to the other older drivers while in actual practice her performance measured via the index was relatively poor. All drivers possessing high leverages were older drivers. Only one driver “O46” had an absolute DFFITS value (-1.42) greater than 1 and there was no absolute value of DFBETAS greater than 1, thus manifesting that driver “O46” was influential. This is also obvious from the Added-variable plot as driver "O46" is quite distant from the rest. since the high leverage cases were not influential therefore, there was nothing wrong with observation of driver "O46" and was left intact as it did not affect the coefficient estimates (i.e. all absolute values of DFBETAS less than 1) (Chatterjee & Hadi, 2006). The outliers i.e. observations having high leverages and high residuals and also the influential cases were scrutinized to ensure that there was no measurement, recording or calculation error. After thorough checking, no error of any sort could be detected manifesting that these were indeed plausible observations. Therefore, the model for the *Optimistic Case* was considered to be representative of all of the observations in the sample and not an artifact of a few.

### 7.4.3 Model Selection Using All Possible Regression Models

The next step was to select a parsimonious model from the full models developed in Section 7.4.1 for each of the scenarios of Section 7.4. Building a model that includes a subset of the available predictors involves two conflicting objectives (Montgomery et al., 2006): (1) The predictors should be as many as possible so that the information contained in them can influence the predicted value (2) since the variance of the prediction $\hat{y}$ and the cost of data collection increases as the number of predictors increase, therefore few predictors are favoured. Deleting predictors has the potential for introducing bias in the estimates of the coefficients and the response, however, if the effect of the deleted-predictors is small, the amount of bias introduced will be less than the reduction in the variance of the response. But if negligible predictors (predictors with zero coefficients or coefficients less than their respective standard errors from the full model) are retained, the variances of the coefficients/parameters and the predicted response will increase (Montgomery et al., 2006). Reduced models will have smaller collinearity if too many predictors are providing the same information (Faraway, 2005).

There are two main types of variable selection procedures: (1) The Stepwise selection approach, which compares successive models, and (2) the Criterion approach that uses optimization of some measure of goodness. Because of the "one-at-a-time" nature of adding/dropping variables, it is possible to miss the "optimal" model in the Stepwise method. Also, because there is so much multiple testing taking place, that the validity of the p-values is dubious i.e., the removal of predictors that are less significant tends to increase the significance of the remaining predictors and in effect results in the overstatement of the importance of the predictors that remain. Whereas, the Criterion based approach typically involves a wider search and compares models in a preferable manner (Faraway, 2005).

The best model was selected via the computation intensive technique of *all possible regression models* and using the criterion approach. In the *all possible regression models* technique, if there are $q$ predictors, then there are $2^q$ models that can be constructed which is based on the fact that each predictor can either be included or excluded from the model. The various candidate models are compared using the criteria that is based on the lack of fit of a model and its *complexity* (Weisberg, 2005). For a candidate subset of predictors, the lack of fit is measured by its SSE (i.e., Error sum of Squares or Residual sum of squares), because as predictors are added to a model SSE decreases. Complexity in the context of multiple linear regression is measured by the number of parameters (including the intercept) included for the candidate subset of predictors in the model. The most common criteria in multiple linear regression and other problems for model comparison is: (1) Akaike Information Criterion (AIC) given by Sakamoto et al. (1987) (2) Bayes Information Criterion (BIC) given by Schwarz (1978), and (3) Mallows' $C_p$ (Mallows, 1973). Smaller values of the entire three criterions are desirable. One important similarity between the three criteria is that if the complexity of the model is fixed (i.e., for example if we are considering all models with say the no. of parameters as four) then all three will agree and the smallest value of each will choose the same model (Weisberg, 2005; Faraway, 2005). AIC and BIC criterion can even be used for comparisons of non-nested models (Cohen et al., 2003; Kass & Raftery, 1995; Chatterjee & Hadi, 2006). Nested means that all the predictors in the smaller model are also present as predictors in the larger model and both models are based on identical observations/cases. BIC and AIC are given by:

$$BIC = n\log_e SSE_p - n\log_e n + p\log_e n \tag{7.5}$$

$$AIC = n\log_e SSE_p - n\log_e n + 2p \tag{7.6}$$

Where,

$n$ = No. of observations.

$SSE_p$ = Error Sum of Squares (Residual Sum of Squares) for the subset regression model.

$p$ = No. of parameters in the model (including constant).

Mallow's $C_p$ is given by:

$$C_p = \frac{SSE_p}{MSE} - (n - 2p) \tag{7.7}$$

Where,

$SSE_p$ = Error Sum of Squares (Residual Sum of Squares) for the subset regression model.

$n$ = No. of observations.

$p$ = No. of parameters in the model (including constant).

$MSE$ = Mean Square Error (i.e., estimate of $\sigma^2$ i.e. variance of error) obtained from the full model.

$C_p$ for the full model will equal $p$. The $C_p$ criterion is used to identify subsets of predictors for which (1) the $C_p$ value is small and (2) $C_p$ value is near $p$ (Kutner et al., 2005). Subsets that possess small values of $C_p$ have small total mean squared error and $C_p$ values near $p$ translates to small bias of the regression model (Kutner et al., 2005). According to Faraway (2005), when the value of $C_p$ is much bigger than $p$ (i.e. considerably above $p$), the model will have a bad fit or a lack of fit (i.e. biased model). The three criteria (i.e. BIC, AIC and $C_p$) trade-off fit of the model in terms of SSE (i.e., Error sum of Squares or Residual sum of squares) against complexity (by providing penalties for adding predictors) (Faraway, 2005).

Using simulation studies, Lutkepohl (Lutkepohl, 1985 cited in Christensen, 2001) showed that the BIC out performed the AIC (and many other criteria) with regard to identification of the correct model and the minimization of prediction errors. Asymptotically, the AIC is not consistent; the BIC has been developed by modifying the AIC to specifically achieve consistency and has out performed the AIC (Christensen, 2001). The BIC is generally considered to solve the model selection problem (Freedman, 2005). The models chosen by BIC tend to be simpler/parsimonious than those chosen by the AIC criterion (Kass & Raftery, 1995) because BIC penalizes larger models more heavily compared to AIC (Faraway, 2005; Kutner et al., 2005). According to Chatterjee and Hadi (2006), the difference between AIC and BIC is in the severity of the penalty for $p$ (i.e. no. of parameters), which is more pronounced in BIC when n > 8, and thus tends to control the over-fitting tendency of AIC. To guard against over-fitting they recommend the use of BIC.

The Coefficient of Multiple Determination ( $R^2$) is defined as a measure of the proportion of variation in the response variable explained by the predictors and is given by:

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}$$

Where SSR is the regression sum of squares, SSE is the error sum of squares (residual sum of squares) and SST is the total sum of squares. $R^2$ is also the square of the correlation between the observed values $y$ and the fitted values $\hat{y}$ (Weisberg, 2005). Adding more predictors to a model increases $R^2$, because SSE can never become larger with more predictors as SST is always the same for the given set of responses. Since $R^2$ can be made larger by including more predictors, an *Adjusted $R^2$* (Adjusted Coefficient of Multiple Determination) is sometimes used that adjusts for the number of predictors in the model. The *Adjusted $R^2$* adjusts $R^2$ by dividing each sum of squares by its respective *degrees of freedom* and is given by (Kutner et al., 2005):

$$\text{Adj-R}^2 = 1 - \frac{\frac{\mathrm{SSE}}{n-p}}{\frac{\mathrm{SST}}{n-1}} = 1 - \left(\frac{n-1}{n-p}\right)\frac{\mathrm{SSE}}{\mathrm{SST}}$$

Where, n is the number of observations and $p$ is the number of parameters in the model (including constant). The adjusted $R^2$ may actually decrease when another predictor is added to the model, because any decrease in SSE may be more than offset by the loss of degrees of freedom in the denominator term *(n-p)* (Kutner et al., 2005).

Since the number of predictors were six, using the technique of *all possible regression models*, $2^6 = 64$ possible models were estimated for each of the three scenarios of Section 7.4. AIC, BIC, $C_p$, $R^2$ and Adjusted $R^2$ were computed for each of the 64 models in each of the scenarios (64×3= 192 models in total) to aid in the decision making process for selection of the best model in each scenario. In each scenario, the following four criteria were used to screen between the 64 models so as to get a smaller subset of models for further scrutiny:

1. The five models having the lowest AIC.
2. The five models having the lowest BIC.
3. The five models having the lowest $C_p$.
4. The one, two, three, four, five, and six predictor model with the lowest AIC, BIC, $C_p$ (Note: all pointed towards the same model in each of the six cases).

A considerable number of common models were screened by the above screening mechanism. In the *Pessimistic Case* scenario, nine models shown in Table 7.8 were screened from the pool of 64 possible models, by using the above four criteria.

**Table 7.8** The nine models screened from the pool of 64 models of *Pessimistic Case* scenario.

| Pessimistic Case (dichotic score = 100 for driver "O56") | | | | | | | |
|---|---|---|---|---|---|---|---|
| Model No. | Predictors | No. of Parameters | $R^2$ | Adj- $R^2$ | Cp | BIC | AIC |
| 1 | $x_2$ | 2 | 0.4129 | 0.4020 | 24.51 | 280.52 | 276.47 |
| 2 | $x_5\ x_6$ | 3 | 0.5337 | 0.5161 | 10.76 | 271.64 | 265.57 |
| 3 | $x_3\ x_1\ x_2$ | 4 | 0.5938 | 0.5703 | 4.94 | 267.95 | 259.85 |
| 4 | $x_1\ x_6\ x_2$ | 4 | 0.5922 | 0.5686 | 5.15 | 268.17 | 260.07 |
| 5 | $x_5\ x_6\ x_2$ | 4 | 0.5765 | 0.5521 | 7.19 | 270.28 | 262.18 |
| 6 | $x_3\ x_1\ x_6\ x_2$ | 5 | 0.6124 | 0.5820 | 4.51 | 269.35 | 259.22 |
| 7 | $x_5\ x_1\ x_6\ x_2$ | 5 | 0.6078 | 0.5770 | 5.11 | 270.01 | 259.88 |
| 8 | $x_4\ x_3\ x_1\ x_6\ x_2$ | 6 | 0.6218 | 0.5840 | 5.28 | 272.00 | 259.84 |
| 9 | $x_4\ x_3\ x_1\ x_6\ x_2\ x_5$ | 7 | 0.6240 | 0.5779 | 7 | 275.70 | 261.52 |

Where,

$\boldsymbol{x_1} = \left(dichotic + 1\right)^{-0.5}$

$\boldsymbol{x_2} = \log_e ufov3$

$\boldsymbol{x_3} = rey - recall$

$\boldsymbol{x_4} = trail$

$\boldsymbol{x_5} = rey - copy$

$\boldsymbol{x_6} = (paper + 1)^{0.5}$

Model No.3 of Table 7.8 had the lowest BIC value (267.95) and a lower value of AIC as well. Its $C_p$ value of 4.94 was small and was near p (No. of parameters). Besides the $R^2$ of model No.3 was 0.5938 as compared to the full model $R^2$ of 0.6240, which shows that by getting rid of half of the predictors, there was a reduction of only 3 percent in the explained variance of driving performance index. The lack of multicollinearity of Model No. 3 was evident from the values of Variance Inflation Factor (VIF), which were 1.35, 1.30 and 1.21. Therefore, Model No.3 was selected as the best and parsimonious model, which was based on the *dichotic*, *ufov3* and *rey-recall* cognitive tests.

In a similar manner, for the *Optimistic Case* scenario and *Deletion of Case of Driver "O56"* scenario, tables showing the screened models from the pool of 64 possible models (in each scenario) are in Appendix-F.

In the *Optimistic Case* scenario, nine models were screened from the pool of 64 possible models. Model No.3 and Model No. 4 of the Table for *Optimistic Case* in Appendix-F had essentially the same and the lowest BIC values (268.81 and 268.86 respectively) and a lower value of AIC as well. The $C_p$ value of both these models was small (5.43 and 5.48 respectively) was small and was near p (No. of parameters). Their $R^2$ values were also essentially the same. Therefore, it did not make much difference which model was selected. It was decided to select model No.4 as this model involved the same predictors as those selected in the *Pessimistic Case*, thereby facilitating comparison. The $R^2$ of Model No.4 was 0.5871 as compared to the full model $R^2$ of 0.6217, which shows that by getting rid of half of the predictors, there was a reduction of only 3 percent in the explained variance of driving performance index. The lack of multicollinearity of Model No. 4 was evident from the values of Variance Inflation Factor (VIF), which were 1.34, 1.30 and 1.19. Therefore, Model No.4 was selected as the best and parsimonious model, which was based on the *dichotic*, *ufov3* and *rey-recall* cognitive tests.

In the *Deletion of Case of Driver "O56"* scenario, ten models were screened from the pool of 64 possible models. Model No.3 of the table for *Deletion of Case of Driver "O56"* in Appendix-F had the lowest BIC value (272.23) and the lowest value of AIC as well. The $C_p$ value of model No.3 was also the lowest (3.57) and was near p (No. of parameters). The $R^2$ of Model No.3 was 0.5227 as compared to the full model $R^2$ of 0.5469, which shows that by getting rid of half of the predictors, there was a reduction of only 2 percent in the explained variance of driving performance index. The lack of multicollinearity of Model No.3 was evident from the values of Variance Inflation Factor (VIF), which were 1.25, 1.23 and 1.12. Therefore, Model No.3 was selected as the best and parsimonious model, which was based on the *dichotic*, *ufov3* and *rey-recall* cognitive tests.

### 7.4.4 Best/Parsimonious Model Development

Since in all three scenarios of section 7.4, the predictors *dichotic*, *ufov3* and *rey-recall* were identified as contributory to best model, it was necessary to determine the correct functional form of model and transformation of predictors for each of the three scenarios when these predictors alone are used in models, as underlined by Montgomery et al. (2006). In developing the best/parsimonious models, all the steps described in Section 7.4.1 relevant to full model development were followed with the exception that there were now three predictors (*dichotic*, *ufov3* and *rey-recall*). The response variable was Tscore_DPI1_weighted. The predictor *dichotic* had a minimum value of zero (i.e. non-positive) for all three scenarios, therefore, a small increment i.e. 1 was added to make it positive. Since the ratio of the maximum to the minimum value of *rey-recall* was less than 10 for all three scenarios, it was decided not to consider it for transformation. Hence, the two variables *dichotic+1* and *ufov3* were considered candidates for transformation in all three scenarios of Section 7.4. The Statistical software R® that is based on the S programming language with the add-on package *alr3* containing the function *bctrans* was used to find the powers $\lambda$ to bring about joint normality using the method of maximum likelihood.

**Table 7.9** Summary table showing estimated $\lambda$ and value adopted for use for the three scenarios for development of the best models.

| Variable | Pessimistic Case | | Optimistic Case | | Deletion of Case of Driver "O56" | |
|---|---|---|---|---|---|---|
| | Estimated $\lambda$ | $\lambda$ rounded for use | Estimated $\lambda$ | $\lambda$ rounded for use | Estimated $\lambda$ | $\lambda$ rounded for use |
| *(dichotic + 1)* | -0.4918 | -0.5 | -0.3944 | -0.5 | -0.3954 | -0.5 |
| *ufov3* | 0.1957 | 0 | 0.1829 | 0 | 0.2505 | 0 |

Table7.9 shows the detail of estimated values of $\lambda$ for each scenario and the value adopted after rounding. After rounding, the adequacy of the rounded values of $\lambda$ was tested through likelihood ratio tests which gave p-values of 0.44, 0.43 and 0.29 for the three scenarios. $\lambda = 0$ is interpreted as a log transformation. Next the three predictors (including two transformed) i.e., $(dichotic + 1)^{-0.5}$ , $\log_e ufov3$ and $rey - recall$ were used in a Box and Cox procedure (Box & Cox 1964 cited in Weisberg, 2005) to find transformation of the response variable Tscore_DPI1_weighted using the statistical software Stata®. Table 7.10 gives values for $\lambda$ along with 95% confidence intervals for the three scenarios for the response variable Tscore_DPI1_weighted .

**Table7.10** Summary table showing estimated $\lambda$ along with 95% confidence intervals for $\lambda$ for the three scenarios for the response variable Tscore_DPI1_weighted for development of best models.

| Pessimistic Case | | Optimistic Case | | Deletion of Case of Driver "O56" | |
|---|---|---|---|---|---|
| Estimated $\lambda$ | 95% confidence interval | Estimated $\lambda$ | 95% confidence interval | Estimated $\lambda$ | 95% confidence interval |
| 2.437 | 0.99 to 3.87 | 2.531 | 1.10 to 3.95 | 2.579 | 1.06 to 4.09 |

Since the confidence intervals for all three scenarios were wide that included two or more of the bench mark levels of $\lambda = -1, -\frac{1}{2}, 0, \frac{1}{2}, 1$ which indicated that $\lambda$ was not crisply estimated. According to Cook and Weisberg (1999), estimates of $\lambda$ outside the range of -2 and +2 often indicate that transformation of the response will not be helpful. Therefore, it was decided to use no transformation for the response variable $\mathrm{Tscore\_DPI1\_weighted}$ for all three scenarios. It may be noted that the same transformations (for the predictors) were arrived at as were used in the Full model development. The following three predictors were regressed against $\mathrm{Tscore\_DPI1\_weighted}$ :

Rey_recall = $rey - recall$

T_dichotic_t_plus = $(dichotic + 1)^{-0.5}$

Ln_ufov3 = $\log_e ufov3$

Results of regression analysis for the *Pessimistic Case* are shown in Table 7.11 to Table 7.13.

**Table 7.11** Detail of Sum of Squares for the scenario *Pessimistic Case* for best model.

| Source | Sum of Squares | Degrees of freedom | MS |
|---|---|---|---|
| Model | 7347.71 | 3 | 2449.237 |
| Residual | 5027.28 | 52 | 96.678 |
| Total | 12375 | 55 | 225 |

**Table 7.12** Detail of F-test and R-squared for the scenario *Pessimistic Case* for best model.

| Detail of F-Test and R-Square | |
|---|---|
| Number of observations | 56 |
| F (3, 52) | 25.33 |
| Prob>F | 0.0000 |
| R-squared | 0.5938 |
| Adj R-squared | 0.5703 |
| Root MSE | 9.8325 |

**Table 7.13** Detail of estimated coefficients and their standard errors for the scenario *Pessimistic Case* for best model.

| Predictors | Coefficient | Standard Error | t | P>\| t \| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| T_dichotic_t_plus | 23.40597 | 7.193504 | 3.25 | 0.002 | 8.971147 | 37.84079 |
| Ln_ufov3 | -7.826133 | 2.042769 | -3.83 | 0.000 | -11.9252 | -3.72701 |
| Rey_recall | 0.7197843 | 0.2512916 | 2.86 | 0.006 | 0.2155308 | 1.224038 |
| constant | 104.3111 | 13.60166 | 7.67 | 0.000 | 77.01737 | 131.6048 |

The test for significance of regression is significant with F (3, 52) = 25.33, p < 0.00005 showing that at least one of the three predictor variable contributes significantly to the model. The individual t-tests of the regression coefficients show all predictors to be highly significant (i.e. all p values way less than 0.05). Results of regression analysis for the *Optimistic Case* and for the *Deletion of Case of Driver "O56"* are in Appendix-F. For the *Optimistic Case*, the test for significance of regression is significant with F (3,52) = 24.65, p < 0.00005 showing that at least one of the predictor variable contributes significantly to the model. The individual t-tests of the regression coefficients show all predictors to be highly significant (i.e. all p values way less than 0.05). For the *Deletion of Case of Driver "O56"*, the test for significance of regression is significant with F (3,51) = 18.61, p < 0.00005 showing that at least one of the predictor variable contributes significantly to the model. The individual t-tests of the regression coefficients show all predictors to be highly significant (i.e. all p values way less than 0.05). The following three best/parsimonious models pertinent to each of the three scenarios were obtained:

- *Pessimistic Case*

$$DPI = 104.31 + 23.40(dichotic + 1)^{-0.5} - 7.83 \log_e(ufov3) + 0.72(rey - recall)$$

- *Optimistic Case*

$$DPI = 105.09 + 22.93(dichotic + 1)^{-0.5} - 8 \log_e(ufov3) + 0.73(rey - recall)$$

- *Deletion of Case of Driver "O56"* scenario

$$DPI = 104.43 + 22.63(dichotic + 1)^{-0.5} - 8.06 \log_e(ufov3) + 0.76(rey - recall)$$

### 7.4.5 Best/Parsimonious Model Diagnostics

After the best/parsimonious models were developed in Section 7.4.4, the next obvious step was to subject them to thorough analysis to ensure that the functional form of the models was correct, the assumptions were satisfied and to identify outliers/influential observations. The diagnostics were performed in a manner similar to that of *Full Model Diagnostics* of Section 7.4.2, therefore the detailed explanations will not be repeated. All graphs and tables relevant to best/parsimonious model diagnostics for each of the three scenarios are at Appendix-F, which can be interpreted as per description of Section 7.4.2. Suffice to say that for all three scenarios, residuals were normally distributed with no evidence of heteroskedasticity or systematic patterns. As a final check, the residuals were also plotted against the square and interaction terms of each of the three predictors to ensure that no systematic pattern was evident (and hence no missing term in the model). Once the functional form of the model was deemed adequate, the next step was to identify and assess outliers and influential observations.

In the *Pessimistic Case*, from the leverages versus squared residuals plot, drivers "Y52", "O6" and "O56" had high leverages (greater than $\frac{2k}{n} = 0.14$) and driver "Y5"' and "O46" had high studentized residuals (2.03 and -3.34 respectively) showing an ill-fit. Based on the three cognitive tests, the value predicted by the model (i.e., y hat) for Driver "Y5" was lower (and hence the positive studentized residual) than her observed value on the index, because her scores on the three cognitive tests were relatively unfavourable compared to the other younger drivers while in actual practice her performance measured via the index was relatively good. The value predicted by the model (i.e., y hat) for Driver "O46" was higher (and hence the negative studentized residual) than her observed value on the index, because her scores on the three cognitive tests were relatively favourable compared to the other older drivers while in actual practice her performance measured via the index was relatively poor. Only one driver "O46" had an absolute DFFITS value (-1.01) greater than 1 and there was no absolute value of DFBETAS greater than 1, thus manifesting that driver "O46" was influential. This is also obvious from the Added-variable plot as driver "O46" is quite distant from the rest. Since the high leverage cases were not influential therefore, there was nothing wrong with observation of driver “O46” and was left intact as it did not affect the coefficient estimates (i.e. all absolute values of DFBETAS less than 1) (Chatterjee & Hadi, 2006). The outliers i.e. observations having high leverages and high residuals and also the influential cases were scrutinized to ensure that there was no measurement, recording or calculation error. After thorough checking, no error of any sort could be detected manifesting that these were indeed plausible observations. Therefore, the model for the *Pessimistic Case* was considered to be representative of all of the observations in the sample and not an artifact of a few.

In the *Optimistic Case*, observations having high leverages and high studentized residuals were identical to that of the *Pessimistic Case* (with slightly different values) and hence the same interpretation. No observation was influential (i.e. all absolute values of DFFITS and DFBETAS less than 1). Therefore, the model for the *Optimistic Case* was considered to be representative of all of the observations in the sample and not an artifact of a few.

In the *Deletion of Case of Driver “O56*” scenario, from the leverages versus squared residuals plot, drivers “Y52” and “O6” had high leverages (greater than $\frac{2k}{n} = 0.14$) and driver “O46” had a high studentized residual (-3.4) showing an ill-fit. Based on the three cognitive tests, the value predicted by the model (i.e., y hat) for Driver “O46” was higher (and hence the negative studentized residual) than her observed value on the index, because her scores on the three cognitive tests were relatively favourable compared to the other older drivers while in actual practice her performance measured via the index was relatively poor. Only one driver “O46” had an absolute DFFITS value (-1.03) greater than 1 and there was no absolute value of DFBETAS greater than 1, thus manifesting that driver “O46” was influential. This is also obvious from the Added-variable plot as driver “O46” is quite distant from the rest. Since the high leverage cases were not influential therefore, there was nothing wrong with observation of driver “O46” and was left intact as it did not affect the coefficient estimates (i.e. all absolute values of DFBETAS less than 1) (Chatterjee & Hadi, 2006). The outliers i.e. observations having high leverages and high residuals and also the influential cases were scrutinized to ensure that there was no measurement, recording or calculation error. After thorough checking, no error of any sort could be detected manifesting that these were indeed plausible observations. Therefore, the model for the *Deletion of Case of Driver "O56"* scenario was considered to be representative of all of the observations in the sample and not an artifact of a few.

### 7.4.6 Model Development Using Cognitive Tests *trail* and *rey-recall*

Since the cognitive tests *trail* and *rey-recall* are relatively quick and easy to administer paper-and-pencil tests, and as they were not together selected as predictors in the best/parsimonious models, it was worthwhile to assess the viability of a multiple linear regression model (for the *Pessimistic Case*) that was based on these two cognitive tests alone. Since the ratio of the maximum to the minimum value of the predictors *trail* and *rey-recall* was less than 10, it was decided not to consider them for transformation (Draper & Smith, 1998; Weisberg, 2005). The response variable *DPI1-weighted* was non-positive and therefore converted into positive values called T-score using Equation 7.2 (as explained in Section 7.4.1) thus obtaining $\text{Tscore\_DPI1\_weighted}$ . Next *trail* and *rey-recall* were used in a Box and Cox procedure (Box & Cox 1964 cited in Weisberg, 2005) to find transformation of the response variable $\text{Tscore\_DPI1\_weighted}$ using the statistical software Stata®. A $\lambda$ value of 2.348 with a 95% confidence intervals of 0.814 to 3.881was obtained. Because of the wide confidence interval and other factors explained in Section 7.4.1, it was decided not to use any transformation on the response $\text{Tscore\_DPI1\_weighted}$ . The predictors *trail* and *rey-recall* were regressed against $\text{Tscore\_DPI1\_weighted}$ . Results of regression analysis are shown in Table 7.14 to Table 7.16.

**Table 7.14** Detail of Sum of Squares when *trail* and *rey-recall* are in the model.

| Source | Sum of Squares | Degrees of freedom | MS |
|---|---|---|---|
| Model | 6349.96 | 2 | 3174.97 |
| Residual | 6025.04 | 53 | 113.68 |
| Total | 12375 | 55 | 225 |

**Table 7.15** Detail of F-test and R-squared when *trail* and *rey-recall* are in the model.

| Detail of F-Test and R-Square | |
|---|---|
| Number of observations | 56 |
| F (2, 53) | 27.93 |
| Prob>F | 0.0000 |
| R-squared | 0.5131 |
| Adjusted R-squared | 0.4948 |
| Root MSE | 10.662 |

**Table 7.16** Detail of estimated coefficients and their standard errors when *trail* and *rey-recall* are in the model.

| Predictors | Coefficient | Standard Error | t | P>\| t \| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| *trail* | -0.2224002 | 0.0489786 | -4.54 | 0.000 | -0.320638 | -0.12416 |
| *rey-recall* | 0.8813328 | 0.266606 | 3.31 | 0.002 | 0.346588 | 1.416077 |
| constant | 98.72119 | 8.214036 | 12.02 | 0.000 | 82.24592 | 115.1965 |

The diagnostics were performed in a manner similar to that of *Full Model Diagnostics* of Section 7.4.2, therefore the detailed explanations will not be repeated. Suffice to say that residuals were normally distributed with no evidence of heteroskedasticity or systematic patterns. As a final check, the residuals were also plotted against the square and interaction terms of each of the two predictors to ensure that no systematic pattern was evident (and hence no missing term in the model). Once the functional form of the model was deemed adequate, the next step was to identify and assess outliers and influential observations.

From the leverages versus squared residuals plot, drivers “O56”, “O39”, “O12” and “Y32” had high leverages (greater than $\frac{2k}{n} = 0.11$) and driver “O46”, “O7” and “O35” had high studentized residuals (-3.371, -2.090 and 2.025 respectively) showing an ill-fit. Based on the two cognitive tests, the value predicted by the model (i.e., y hat) for Driver “O35” was lower (and hence the positive studentized residual) than her observed value on the index, because her scores on the two cognitive tests were relatively unfavourable compared to the other older drivers while in actual practice her performance measured via the index was relatively good. The value predicted by the model (i.e., y hat) for Driver "O46" and "O7" was higher (and hence the negative studentized residuals) than her observed value on the index, because their scores on the two cognitive tests were relatively favourable compared to the other older drivers while in actual practice their performance measured via the index was relatively poor. No observation was influential (i.e. all absolute values of DFFITS and DFBETAS less than 1). Lack of multicollinearity was evident from the values of Variance Inflation Factors (VIF), which were 1.25 and 1.25. The outliers i.e. observations having high leverages and high residuals were scrutinized to ensure that there was no measurement, recording or calculation error. After thorough checking, no error of any sort could be detected manifesting that these were indeed plausible observations. Therefore, the model with the predictors *trail* and *rey-recall* was considered to be representative of all of the observations in the sample and not an artifact of a few.

### 7.4.7 Validation of Models

The best/parsimonious models developed in Section 7.4.4 and the model based on the two predictors (*trail* and *rey-recall*) of Section 7.4.6 needed to be validated. Validation of regression models manifests the reasonableness and stability of regression coefficients, the utility and plausibility of the regression function and the ability to generalize inferences that are drawn from the model (Kutner et al., 2005). The basis for many of the techniques that are employed in model development is the fit of the model to the available data, and therefore it apparently may seem that a well fitting model will also be a successful predictor; however, this is not necessarily the case. Therefore, formal techniques of model validation have to be employed. The three fundamental techniques of validating a regression model are (Montgomery et al., 2006):

- Investigate the model's predictive ability by collecting new (or fresh) data.

- Use a holdout sample to check the model and its predictive performance (data splitting i.e., setting aside some of the original data sample). Some times, data splitting is called cross validation.
- Comparison of results with prior experience, theoretical expectations, empirical results and simulation results etc.

The best validation technique is through the collection of new data (Kutner et al., 2005), and then to compare the predictions made by the model against them. Accurate predictions of new data by the model will instill confidence in the model and the model-building process. However, collection of new data is not always practical or feasible because of limited resources, dismantled experimental plant etc. In such situations, the holdout technique (data-splitting) holds more promise, when the data set is large. In the data-splitting technique, the data set is split into two sets, the first set is called the *model-building set* or the *training sample* and is used to develop the model, while the second set is called the *validation* or *prediction set* and is used to evaluate the reasonableness and predictive ability of the model. The regression model is developed from the *model-building set* and the *validation set* is then used to evaluate the predictive ability of the model. One potential drawback of the data splitting method is that it reduces the precision with which the regression coefficients are estimated (Kutner et al., 2005). That is, the standard errors of the regression coefficients that are estimated from the *model-building set* will be larger than they would have been if all the data (i.e., without splitting) had been used up in estimating the regression coefficients (Kutner et al., 2005; Montgomery et al., 2006; Snee, 1977). However, if the *model-building set* is reasonably large, then the standard errors will in general not be much larger than they would have been if all the data (i.e., without splitting) had been used in estimating the regression coefficients.

Since our sample was relatively small, using the data-splitting technique of model validation would have resulted in imprecise regression estimates from the *model-building set* thereby resulting in an unreliable model (to be validated) which would have negated the whole exercise/essence of model validation.

Harrell et al. (1996) recommend the use of the entire sample in model development as according to the authors, data are too precious to waste. However, refinements of data-splitting exist. For smaller data sets, the *K-fold cross-validation* procedure is employed

(Kutner et al., 2005). In the *K-fold cross-validation* procedure, the data are split into $K$ roughly equal parts. Out of the $k$ = 1, 2, ….., $K$ parts, the $k^{th}$ part is used as the *validation set* and the remaining $K$-1 parts are used as the *model-building set* and a measure of prediction error is obtained; the $K$ estimates of prediction errors are then combined to produce a *K-fold cross-validation estimate*. When the data split is such that $K= n$, it becomes the PRESS statistic variant of data-splitting (Montgomery et al., 2006; Weisberg, 2005; Snee, 1977). In the PRESS Statistic variant of data-splitting, almost all the data (except one observation) is used to estimate the regression coefficients/model. A model is estimated from all observations except observation $i$ . Then this model is used to predict the fitted value $\hat{y}_{(i)}$ of observation $i$. Thus a model is used to predict a response variable for an observation whose data was not used in the development of the model. The *prediction error* of observation $i$ is given by:

$$e_{(i)} = y_i - \hat{y}_{(i)} \tag{7.8}$$

Where,

$e_{(i)}$ = prediction error for observation $i$.

$y_i$ = Observed value of observation $i$.

$\hat{y}_{(i)}$ = Fitted value from model estimated from all observations except $i$.

This *prediction error* calculation is repeated for all observations $i$ = 1, 2,…, $n$. These *prediction errors* are called PRESS residuals. It is possible to calculate $n$ PRESS residuals without fitting $n$ different regressions. The $i^{th}$ PRESS residual is given by:

$$e_{(i)} = \frac{e_i}{1-h_{ii}} \tag{7.9}$$

Where,

$e_{(i)}$ = prediction error for observation $i$ ($i^{th}$ PRESS residual).

$e_i$ = Ordinary residual of observation $i$.

$h_{ii}$ = $i$ th diagonal element of the hat matrix.

Large PRESS residuals have the potential to identify observations where the model does not fit the data well or observations where the model is likely to provide poor future predictions (Montgomery et al., 2006). The *prediction error* sum of squares (i.e., the sum of squared PRESS residuals) called the PRESS statistic is used as measure of model quality. The PRESS statistic is:

$$\text{PRESS} = \sum_{i=1}^{n} \left[ y_i - \hat{y}_{(i)} \right]^2 \tag{7.10}$$

$$= \sum_{i=1}^{n} \left( \frac{e_i}{1 - h_{ii}} \right)^2$$

A model with a small value of PRESS is favoured. PRESS is considered as a measure of how well a model performs in predicting new data (Montgomery et al., 2006). It is used to calculate an $R^2$-like statistic ($R^2_{\text{prediction}}$) for prediction, which gives an indication of the predictive capability of the model:

$$R^2_{\text{prediction}} = 1 - \frac{\text{PRESS}}{SS_T} \tag{7.11}$$

Where, $SS_T$ is the total sum of squares. The equation for $R^2_{\text{prediction}}$ is analogous to that of the Coefficient of Multiple Determination $R^2$ given by:

$$R^2 = 1 - \frac{SSE}{SS_T} \tag{7.12}$$

$SSE$ is the error sum of squares given by $\sum_{i=1}^{n} \left( y_i - \hat{y}_i \right)^2$, where $y_i$ is the observed value of observation $i$ and $\hat{y}_i$ is the fitted value of observation *i*. Table 7.17 shows the PRESS statistic, $R^2$ and $R^2_{\text{prediction}}$ for the best/parsimonious models of the three scenarios developed in Section 7.4.4 and the model developed for the two predictors (*trail* and *rey-recall*) in Section 7.4.6.

**Table 7.17** PRESS statistic, $R^2$, and $R^2_{\text{prediction}}$ for the four models developed in Sections 7.4.4 and 7.4.6.

| Regression Model | PRESS statistic | $R^2$ | $R^2_{\text{prediction}}$ |
|---|---|---|---|
| *Pessimistic Case* | 5831.133 | 0.59 | 0.53 |
| *Optimistic Case* | 5933.261 | 0.59 | 0.52 |
| *Deletion of Case of Driver "O56"* | 6701.847 | 0.52 | 0.45 |
| Two predictor model (*trail* and *rey-recall*) | 6744.327 | 0.51 | 0.46 |

Table 7.17 shows for example that for the *Pessimistic Case* model, $R^2_{prediction}$ was 0.53 and $R^2$ was 0.59 thus it is expected of this model to explain 53 percent of the variability of driving performance index in new data as compared to 59 percent of the variability explained by the least-squares fit to the *model-building* data set. The loss/degradation in $R^2$ with regard to the four models of Table 7.17 for the *Pessimistic Case*, *Optimistic Case*, *Deletion of Case of Driver "O56"* and the *Two predictor model* was 0.06, 0.07, 0.07 and 0.05 respectively. The loss/degradation in $R^2$ is small keeping in view that Montgomery et al. (2006) have termed a loss/degradation of 0.0632 (i.e. 6.32 percent) as slight and Ratrout et al. (2004) have made the following comment relevnt to a degradation in $R^2$ of 0.128 (i.e. 12.8 percent) when validating their model: "Nevertheless, the degradation in $R^2$ is not severe enough to suspect a serious inconsistency in the behavior of Model 1". Therefore, there is reasonably strong evidence that the models will be satisfactory predictors. It would have been useful to have an accepted statistical test to judge the acceptable degradation in $R^2$; however, such a test is not available.

The signs of the coefficients of all four models of table 7.17 were correct and their magnitudes reasonable. The variance inflation factors (VIF), which is a multicollinearity diagnostic also serves as an important guides to the validity of a model. If a VIF of a predictor exceeds 5 or 10 (Montgomery et al., 2006; Snee, 1977), it indicates that its coefficient is poorly estimated or unstable due to the near-linear dependencies that exist among the predictors / regressors. Variance Inflation Factors(VIFs) relevant to all predictors in all four models were way down below 5 thereby manifesting stability of regression coefficients.

## 7.4.8 Results and Discussion

The best/parsimonious models developed for the three scenarios of Section 7.4 are given below in Equations 7.13, 7.14 and 7.15:

- *Pessimistic Case*

$$DPI = 104.31 + 23.40(dichotic + 1)^{-0.5} - 7.83 \log_e(ufov3) + 0.72(rey - recall) \tag{7.13}$$

- *Optimistic Case*

$$DPI = 105.09 + 22.93(dichotic + 1)^{-0.5} - 8\ \log_e(ufov3) + 0.73(rey\text{ - }recall) \tag{7.14}$$

- *Deletion of Case of Driver "O56"* scenario

$$DPI = 104.43 + 22.63(dichotic + 1)^{-0.5} - 8.06\ \log_e(ufov3) + 0.76(rey\text{ - }recall) \tag{7.15}$$

The model based on the two predictors (*trail* and *rey-recall*) of Section 7.4.6 is given below by Equation 7.16:

$$DPI = 98.72 - 0.22(trail) + 0.88(rey - recall) \tag{7.16}$$

The regression models for the three scenarios (i.e. Equations 7.13, 7.14 and 7.15) are essentially the same as they have the same functional form, the same predictors, essentially the same coefficients and essentially the same predictive performance. Therefore, in practice it will make little difference which of these is adopted. This shows that the single missing value of driver "O56" for the dichotic listening test did not have much bearing on the development of the models. However, it was decided to adopt the model relevant to the *Pessimistic Case* (Equation 7.13) as it had the lowest BIC value (267.95) and a higher $R^2$ value (0.59). The Coefficient of Multiple Determination ($R^2$) is a global statistic used to assess the fit of a model. The relatively high value of 0.59 showed that 59 percent of the variation in the Driving Performance Index (DPI) was explained by the three cognitive tests. The interpretation for the partial regression coefficient of *rey-recall* is that a unit increase in *rey-recall* would result in an increase of 0.73 in the Driving Performance Index (DPI) when all other predictors (i.e. *dichotic* and *ufov3*) are held constant. Partial regression coefficients relevant to *ufov3* and *dichotic* can not be interpreted as these predictors are in transformed form. The mismatch (as evident from the not-too-high $R^2$) between the three neuropsychological tests and Driving Performance Index (DPI) is because besides cognitive functions, the role of experience, adaptability and motivation can not be downplayed in driving (Ranney, 1994; Lundqvist, 2001; Lundqvist & Ronnberg, 2001 cited in Radford et al., 2004). The model based on *trail* and *rey-recall* could not compete with the *Pessimistic Case* model (Equation 7.13) as it had a lower $R^2$ value (0.51) and a higher BIC value (274.06). Although these two tests (*trail* and *rey-recall*) are paper-and-pencil tests that are easy to administer, together they failed to account for a higher percentage of variability in Driving Performance Index (DPI) than the other three tests (*dichotic, ufov3* and *rey-recall*) as a group. One potential draw-back of especially using the *trail* test for such predictive purposes could be that a candidate could get hold of a standard testing sheet and through practice acquaint himself with the spatial configurations of letters and alphabets which would allow him to get a higher but biased score. The Driving Performance Index (DPI) in these models may be regarded as a general driving performance index that was obtained by considering all 24 driving performance parameters. This index gives an idea of the *general driving skill* of a driver with essentially the same emphasis being placed on each driving performance parameter, be it for example the *No. of Total Hazards*, *Over Speed Limit (Percent of Time)* or *Standard Deviation in Speed DA Task* and therefore, cannot be used in the identification of drivers exhibiting risky driving behavior due to cognitive impairment. In order to identify drivers exhibiting risky driving behavior due to cognitive impairment, it is necessary that the effects of parameters that assess driving skills at the "controlled processing level" ("effortful" processing) be isolated from the rest of the parameters, before undertaking evaluation of risky-driving behavior based on these broader categories of parameters. Since these effects were not isolated in the Driving Performance Index (DPI) in the model, therefore this index is not capable of discriminating between risky driving behavior due to cognitive impairment and normal driving behaviour.

Although, the model involving the *trail* test did not stand out higher with regard to explaining the variation in the Driving performance Index (DPI), however, the *trail* test has correlated strongly with a wide range of driving performance measures in participants with Parkinson's Disease (Stolwyk et al., 2006); McKenna (McKenna, 1998) had some success in using it to screen drivers; has high sensitivity to the progressive cognitive declines that take place during the course of Dementia (Storandt et al., 1984 cited in Stutts et. al., 1998); is recommended by the AMA (The American Medical Association) due to its correlation with driving performance (in simulators and on-road tests) in older adults *with dementia* (Carr et al., 2006). As obvious from the cited studies, the *trail* test is sensitive when a state of neurologic disease exists (or the population is a clinical population), in these situations it is bound to show some good correlation with driving performance / screening because of covariability. Our sample consisted of active drivers from the general driving population (not a clinical population), who apparently had good mental and physical constitution. Therefore, the *trail* test was bound to perform less favourably in gauging driving skill. According to Carr et al. (2006), a score of 180 seconds or longer on the *trail* test (i.e. Trail Making-B test) indicates an increased risk of unsafe driving. Based on normal mixture cluster analysis, out of 56 drivers, 8 drivers were identified as possessing poor driving skills, whereas out of these 8 drivers, only one driver ("O56") had a *trail* score of 180 seconds or greater (his score was 191 seconds). This clearly shows the tendency of the *trail* test to identify only the most severe cases (which are relatively easy to identify) and its failure to demarcate the drivers/cases having doubtful skills (which are the most challenging to identify for further on-road tests). The cognitive tests *dichotic*, *ufov3* and *rey-recall* which were identified as predictors in the best/parsimonious model were also intuitively the most appropriate tests bearing relevance for driving skill. Reger et al.(2004) conducted meta-analysis of 27 primary studies to examine the relationship between neuropsychological functioning and driving ability for adults with dementia and concluded that visuospatial skills and attention were most helpful in screening at risk drivers. Christie et al. (2001a) identified visual memory, executive abilities, spatial awareness and attention as predictors of driving skill. The three cognitive tests (*dichotic*, *ufov3* and *rey-recall*) as a whole measure visual-spatial construction ability, visual memory, organizational, planning and problem solving skills (executive functioning), motor functioning, ability to switch attention, vulnerability to distraction, selective attention, visual processing speed, divided and spatial attention. Impairments in visual memory (which are primarily measured by *rey-recall*) also contributes to poor decision making (based on incorrect input) thereby increasing drivers' risk of errors, crashes and injuries (Rizzo & Kellison, 2004). Kahneman et al. (1973) reported correlation coefficients ranging from 0.29 to 0.37 between errors on the dichotic listening test and crashes over a one year period in young professional bus drivers aged 22 to 32. Parasuraman & Nestor (1991) give a list of other studies (along with relevant data in tabular format) that show significant correlation starting from 0.3 to 0.4 between measures of driving performance (using accident rate or closed-course driving performance index) and at least one measure of selective attention (which is measured by the dichotic listening test). Parasuraman & Nestor (1993) also report that the ability to switch attention (such as in the dichotic listening test) is also correlated with driving performance in normal individuals. Numerous studies have been cited in Section 3.7 which show a significant relationship between the UFOV test and driving performance/accidents.

## 7.5 Summary

This chapter has described the development of linear models, their diagnostics and validation. The driving performance index DPI-weighted was regressed against neuropsychological tests. The rationale and procedure for the initial screening of the nine candidate predictors (cognitive tests) to six predictors (*trail*, *rey-copy*, *rey-recall*, *dichotic*, *paper* and *ufov3*) has been highlighted. The methodology adopted to resolve the issue of the missing value of driver "O56" (a vital case) on the dichotic listening test has been described considering three scenarios:

- Assigning the most optimistic outcome to the missing value (Optimistic case).
- Assigning the most pessimistic outcome to the missing value (Pessimistic case).
- Dropping the missing value observation/case (*Deletion of Case of Driver "O56"* scenario).

The process of development of full models for the three scenarios has been explained in detail along with diagnostics through the aid of statistical tests and graphical displays including the identification of outlier and influential points. Best/parsimonious model selection using the computation intensive technique of *all regression models* along with the AIC, BIC and $C_p$ criterion as measure of goodness, has been demonstrated for all three scenarios. The process of best/parsimonious model development and their diagnostics through the aid of statistical tests and graphical displays including the identification of outlier and influential points has been elaborated upon. A separate model has been developed for two of the easy to administer paper-and-pencil tests (*trail* and *rey-recall*). Validation methodology has been outlined and all models validated with reasonably strong evidence that they will be satisfactory predictors. The best/parsimonious regression models for the three scenarios are essentially the same as they have the same functional form, the same predictors, essentially the same coefficients and the same predictive performance. Therefore, in practice it will make little difference which of these is adopted. In the model, 59 percent of the variation in the Driving Performance Index (DPI), which can be regarded as an index of *general* driving ability, is explained by the *dichotic, ufov3* and *rey-recall* cognitive tests. The model based on *trail* and *rey-recall* can not compete with this model, on account of its relatively low $R^2$ value and relatively higher BIC value. Logistic regression modeling using the dichotomous categorization of drivers as the dependent variable and cognitive tests as predictors is described in the next chapter.

# 8 Logistic Regression Model

## 8.1 Introduction

The objectives of this chapter were to develop logistic regression models that would discriminate "poor-drivers" from "not-poor-drivers" by using: (1) A single cognitive test that brings about the best discrimination (2) Age as predictor (3) A composite measure of all nine cognitive tests as a predictor. And the identification of the model that brings about the best discrimination amongst the three models. To model the "poor-drivers" through logistic regression using different cognitive tests, it was necessary that the dependent variable (driver grouping) be dichotomized into "poor drivers" and "not-poor-drivers" from the three-group classification obtained in Sections 6.4.7 and 6.5. In Sections 6.4.7 and 6.5, a clustering regime based on the variables *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* and three clusters was found to provide the best solution. The three groupings of drivers and their mean values on the three variables are reproduced in Table 8.1.

**Table 8.1** Mean values of *DPI3-weighted, No. of Total Hazards* and *No. of Low-speed Warnings* for model EEV with BIC of -1075.632.

| Groups for model EEV with 3 clusters/groups having BIC of -1075.632 | | | | |
|---|---|---|---|---|
| Group No. | Number of Drivers | Mean *No. of Total Hazards* | Mean *No. of Low-Speed warnings* | Mean *DPI3-weighted* |
| 1 | 20 | 8.1 | 441.55 | -0.2607223 |
| 2 | 8 | 16.5 | 553 | -0.8960832 |
| 3 | 28 | 7.321429 | 381.0357 | 0.4422541 |
| Maximum & Minimum value of *No. of Total Hazards* = 30, 2<br>Maximum & Minimum value of *No. of Low-Speed warnings* = 656, 286<br>Maximum & Minimum value of *DPI3-weighted* = 1.158719, - 2.083781 | | | | |

On average, group No.3 had the highest score on *DPI3-weighted* (a driving index), the lowest *No. of Total Hazards* and the lowest *No. of Low-speed Warnings*. Similarly, group No.1 had favourable scores on all these variables compared with group no. 2. Despite the fact that drivers in Group no.2 were on average driving at the lowest speeds, they had the greatest number of accidents etc (i.e. no. of total hazards) and had low rating on all other driving performance measures (i.e. *DPI3-weighted*). This substantiates their classification as “poor drivers”. Figure 8.1 shows a driver-classification graph of *No. of Low-speed Warnings* against *No. of Total Hazards* for the driver groupings of Table 8.1. The square symbols represent group No.2, the triangular symbols group No.1, and the circular symbols represent group No.3 (the ellipses superimposed on the classification plot correspond to the multivariate analogs of the standard deviations for each mixture component (i.e. they correspond to the covariances of the components) with centers at the means ). Table 8.1 and Figure 8.1 indicate that the attributes of group No. 2 are quite deviant from the other two groups (i.e. group No. 1 & 3), which are relatively quite close. Since in logistic regression the dependent variable must be dichotomous, it was logical to merge group No. 1 and 3 to form a “not-poor-drivers” group and designate group No. 2 as the “poor-drivers” group.

Thus the “poor drivers” group consisted of 8 drivers and the “not-poor-drivers” group of 48. Following the conventions of logistic regression and our objective, the 8 drivers were coded as 1 (i.e. the event of interest occurred) and the remaining 48 as zero (i.e. the event of interest did not occur) under the dependent variable category. To predict the binary response (i.e. “poor drivers” and “not-poor-drivers”) through different cognitive tests, logistical modeling was chosen instead of the alternative technique of *discriminant analysis* because in *discriminant analysis* the assumption that the predictors follow a joint multivariate normal distribution is rarely if ever satisfied (Kutner et al., 2005; Hosmer and Lemeshow, 2000). Besides, logistic regression modeling is remarkably flexible and unless the data set has most of the probabilities as very small or very large, or where the fit is extremely poor (which can be identified systematically), it is unlikely for any other modeling technique to provide a better fit (Hosmer and Lemeshow, 2000). There are no assumptions required regarding the distributions of predictors in logistic regression (Tabachnick & Fidell, 2007). The following sections describe the details of the development of logistic regression models.

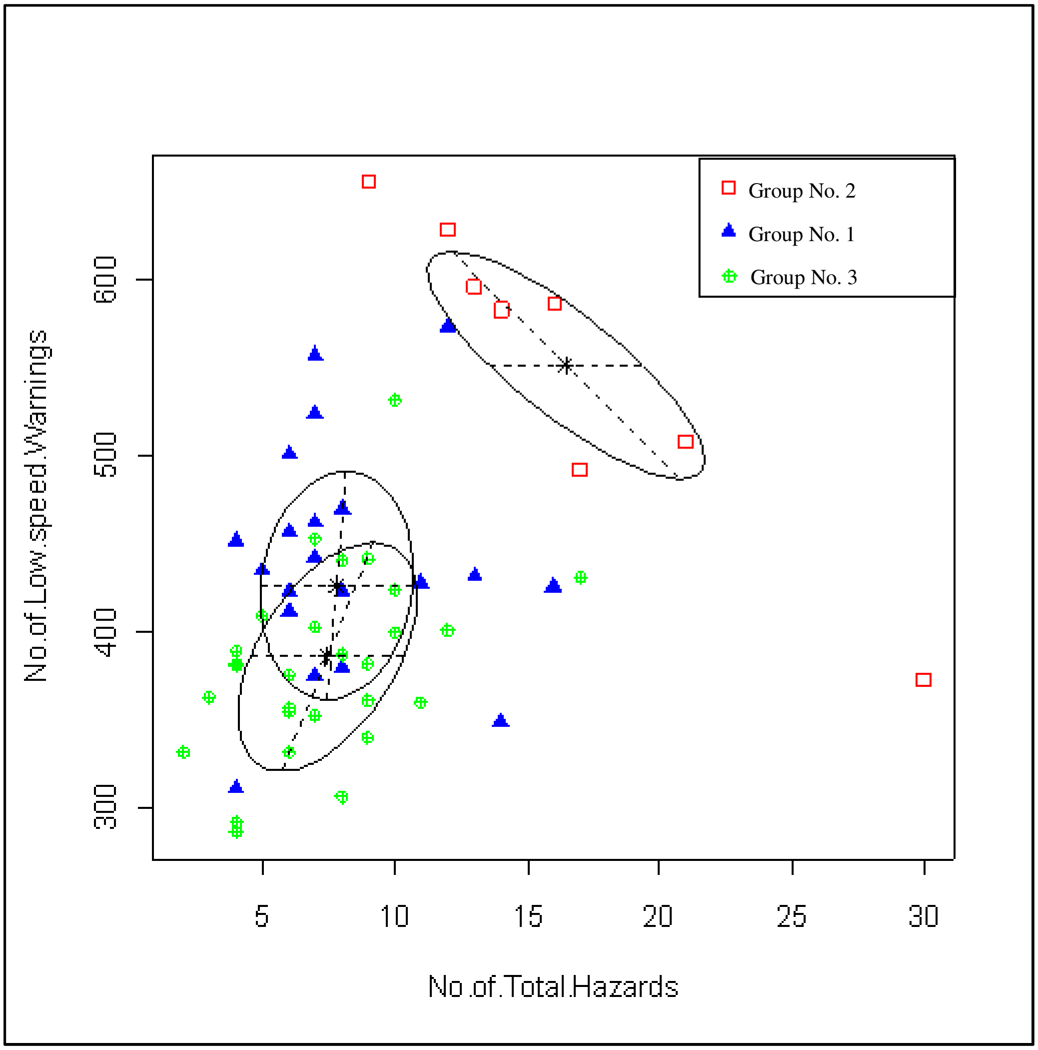



**Figure 8.1** Driver-classification graph of *No. of Low-speed Warnings* against *No. of Total Hazards* for model EEV with 3 groups/clusters having BIC of -1075.632.

## 8.2 Sample size considerations

In logistic regression, the quite general approach of ML (Maximum Likelihood) is used in estimation primarily because of two reasons (Allison, 1995). The estimators produced by ML have good large-sample properties. ML estimators are consistent, asymptotically efficient, and asymptotically normal provided that certain regularity conditions are met. To clarify further, Allison (1995) elaborates that *Consistency* means that as the sample becomes larger, the estimates converge in *probability* to the true values (i.e. the estimates will be approximately unbiased in large samples). *Asymptotically efficient* means that the estimates will have standard errors that are (approximately) at least as small as that for any other estimation method, in large samples. And *Asymptotically normal* means that in large samples, the sampling distribution of the estimates will be approximately normal, and therefore the normal and chi-square distributions can be used to compute confidence intervals and *p*-values. All these approximations get better as sample size gets larger. According to Allison (1995), both for large samples and small, researchers routinely use ML estimation, but he urges caution in interpreting *p*-values and confidence intervals when the samples are small and considers the use of smaller *p*-values to be more reasonable in order to compensate for the poor approximation to the normal and chi-square distributions (for the likelihood ratio and Wald tests), in small samples.

The indications of problems of small number of observations/cases relevant to the number of predictors are that extremely high parameter estimates and standard errors result (Tabachnick & Fidell, 2007); the natural course of action then would be to either increase the number of observations or eliminate one or more predictors.

The number of parameters (of logistic regression) that can be reasonably estimated with maximum likelihood depends not only on the overall sample size but on factors such as the marginal balance on the "1's" and "zeros" of the dependent variable (i.e. the degree of lop sidedness of the dependent variable; in other words when there are few 1's and lots of "zeros", or vice versa) and the distribution of the predictors (Jewell, 2004). In logistic regression, there should be about 10 events (the number of events being the smaller of the number of "1's" and "zeros" of the dependent variable) per predictor variable in order to obtain reasonably stable estimates of regression coefficients (Belle, 2002; Jewell, 2004) and parameters no more than 10 percent of the sample size (Jewell, 2004). This is because, in order to detect partial effects, multiple logistic regression requires larger sample sizes (Agresti, 2007).

Keeping in view the overall sample size, the number of candidate predictors , and the lop sidedness of the dependent variable (i.e. 8 versus 48), the decision to begin a multivariate logistic model based on all possible predictors (*trail, clock, rey-copy, rey-recall, dichotic, paper, ufov1, ufov2, ufov3*) or a smaller subset of them was not appropriate as it would have resulted in a numerically unstable model (thereby making use of the likelihood ratio tests unreliable). Instead, as per recommendation of Hosmer and Lemeshow (2000) it was decided to fit nine univariate models, one for each predictor (i.e. each cognitive test) and use a variety of criterion to aid decision making in selecting the best model.

## 8.3 Selecting the Best Univariate Model

Because of the missing value encountered for the dichotic listening test with driver "O56", univariate logistic regression models were developed for the following three scenarios using the statistical software Stata®:

1. Pessimistic Case (dichotic score = 100 for driver "O56").
2. Optimistic Case (dichotic score = 17.11 for driver "O56").
3. Deletion of Case of Driver "O56".

The dependent variable y was coded as 1 for "poor-drivers" and as 0 for "not-poor-drivers" and the nine cognitive tests were considered one at a time for each of the aforementioned three scenarios. Details of the technique of logistic regression can be found in Hosmer and Lemeshow (2000), Agresti (2007), Kutner et al. (2005), Long (1997) and Long and Freese (2006). A number of Goodness-of-fit statistics relevant to each model were computed and are shown in Table 8.2 for the *Pessimistic Case* scenario.

**Table 8.2** Goodness-of-fit statistics for univariate logistic regression models for the *Pessimistic Case* scenario.

| Univariate Logistic Regression for *Pessimistic case* (dichotic score=100) | | | | | |
|---|---|---|---|---|---|
| Univariate Model with Predictor | *P*-value | Deviance | BIC | AIC | Pseudo-$R^2$ |
| *Ufov3* | 0.0001 | 29.68 | 37.73 | 33.68 | 0.3539 |
| *Dichotic* | 0.0002 | 32.07 | 40.12 | 36.07 | 0.3017 |
| *Trail* | 0.0015 | 35.85 | 43.9 | 39.85 | 0.2195 |
| *Rey-copy* | 0.0032 | 37.27 | 45.32 | 41.27 | 0.1886 |

| Univariate Logistic Regression for *Pessimistic case* (dichotic score=100) | | | | | |
|---|---|---|---|---|---|
| Univariate Model with Predictor | *P*-value | Deviance | BIC | AIC | Pseudo-$R^2$ |
| *Paper* | 0.0048 | 37.98 | 46.03 | 41.98 | 0.1731 |
| *Clock* | 0.0069 | 38.64 | 46.69 | 42.64 | 0.1587 |
| *Rey-recall* | 0.0089 | 39.08 | 47.13 | 43.08 | 0.1492 |
| *Ufov2* | 0.0091 | 39.14 | 47.19 | 43.14 | 0.148 |
| *Ufov1* | 0.2147 | 44.39 | 52.44 | 48.39 | 0.0335 |

The P-value in Table 8.2 is relevant to the likelihood ratio test and is based on the comparison of the Full and Reduced models, where the Full model is the current model and in the present context includes the constant $\beta_0$ and $\beta_1$ and the Reduced Model consists of the null model i.e. a model based on only the constant $\beta_0$. In other words, $H_0$: $\beta_1 = 0$ against $H_a$: $\beta_1 \neq 0$. It is an omnibus test to see if the model as a whole is statistically significant. The test statistic is given by:

$$G^2 = -2\left[\log_e L(R) - \log_e (F)\right] \tag{8.1}$$

Where, $\log_e L(R)$ is the log likelihood of the Reduced Model (null model in this case) and $\log_e (F)$ is the log likelihood of the Full Model (model with intercept and predictor). When $H_0$ holds, $G^2$ has a $\chi^2$ distribution with degrees of freedom equal to the difference in the degrees of freedom of the Full and Reduced model (i.e. difference in the number of parameters of the two models). To illustrate the case when only the predictor *ufov3* is in the model (see Table 8.2) we get (log likelihoods in Appendix-G):

$$G^2 = -2\left[-22.966514 - (-14.839488)\right] = 16.25 \tag{8.2}$$

Since $G^2$ = 16.25 with 1 degree of freedom (D.F) has a P-value of 0.0001, we conclude $H_a$ that ufov3 is highly significant and should not be dropped from the model.

The Deviance compares a given model with a saturated model. A saturated model has as many parameters as there are data points, so the data is reproduced by the model perfectly. The Deviance statistic does not have a chi-squared distribution and is given by:

$$D = -2\left[\log_e L(\text{fitted model}) - \log_e L(\text{saturated model})\right] \tag{8.3}$$

The Deviance in context of logistic regression plays the same role as the error sum of squares (residual sum of squares) plays in linear regression and has a prominent role in some approaches in assessing goodness-of-fit (Hosmer and Lemeshow, 2000). $\text{Log}_e\, L$ (fitted model) is the log Likelihood of the current model. In logistic regression, the Likelihood of the saturated model is 1 (and therefore Log Likelihood is zero). To illustrate the case when only the predictor *ufov3* is in the model (see Table 8.2) we get (log likelihoods in Appendix-G):

$$D = -2(-14.839488) = 29.68 \tag{8.4}$$

BIC (Bayesian information criterion) and AIC (Akaike's information criterion) is given by (Kutner et al., 2005):

$$BIC = -2\log_e L(b) + p\log_e(n) \tag{8.5}$$

$$AIC = -2\log_e L(b) + 2p \tag{8.6}$$

Where $\log_e L(b)$ is the log Likelihood of the fitted model, $p$ is the number of parameters in the model (including constant) and n is the number of observations. Promising models will have relatively small values of BIC and AIC. The criterion is used to assess the overall fit of a model and can be used to compare both nested and non-nested models, however, the BIC is gaining more popularity (Long, 1997). To illustrate the case when only the predictor *ufov3* is in the model (see Table 8.2) we get (log likelihoods in Appendix-G):

$$BIC = -2(-14.839488) + 2\ \log_e 56 = 37.73 \tag{8.7}$$

$$AIC = -2(-14.839488) + 2(2) = 33.68 \tag{8.8}$$

The Pseudo-$R^2$ (also called McFadden's $R^2$) compares a model with just the intercept (i.e. null model) to a model with other parameters besides the intercept (Long and Freese, 2006). It is the proportion of change in terms of the Log Likelihood and is given by:

$$R^2_{\text{McF}} = \frac{\log_e(\text{Intercept}) - \log_e(\text{Full})}{\log_e(\text{Intercept})} = 1 - \frac{\log_e(\text{Full})}{\log_e(\text{Intercept})} \tag{8.9}$$

Where $\log_e$ (Full) is the Log likelihood of the full model (i.e. model with other parameters besides the intercept) and $\log_e$ (intercept) is the Log Likelihood of the intercept only model (i.e. the null model). Analogous to $R^2$ in linear regression, $\log_e$ (Intercept) is thought of as the total sum of squares and $\log_e$ (Full) as residual sum of squares (i.e. error sum of squares) (Long, 1997). Pseudo-$R^2$ always increases as variables are added to the model. Pseudo-$R^2$ values are low compared to typical $R^2$ values that are encountered in good linear regression models, because low Pseudo-$R^2$ values in logistic regression are the norm. However, as a statistic in evaluating competing models, they may be helpful in the model building stage (Hosmer and Lemeshow, 2000). To illustrate the case when only the predictor *ufov3* is in the model (see Table 8.2) we get (log likelihoods in Appendix-G):

$$R^2_{\text{McF}} = 1 - \frac{-14.839488}{-22.966514} = 0.3539 \tag{8.10}$$

Both the AIC and BIC involve penalties that are based on the number of parameters in the model. Therefore, both criteria trade-off fit of the model in terms of Log Likelihood against complexity (by providing the penalties $p \log_e(n)$ and $2p$ for adding predictors in the BIC and AIC expressions respectively). A major draw-back of using deviance as a model selection criterion is that the Deviance (which essentially is $-2 \log_e L(b)$) will never increase as terms are added to the model because there is no penalty for adding predictors.

The best model for the *Pessimistic Case* scenario evident from Table 8.2 is a univariate model based on the *ufov3* cognitive test because of the lowest BIC, AIC and even Deviance. The Pseudo-$R^2$ of this model is also the highest and the P-value (0.0001) for the likelihood ratio test indicates that the model is statistically highly significant. The next best univariate models in decreasing order of over all fit are the models based on *dichotic*, *trail*, *rey-copy* and *paper* cognitive tests (ancillary detail at Appendix-G). This procedure was repeated for the *Optimistic Case* scenario and the *Deletion of Case of Driver "O56"* scenario and the same results were obtained (detail at Appendix-G). The best model was based on the *ufov3* test, and since there was no missing value present on this test, it was decided to develop a single univariate model based on all 56 observations of the *ufov3* test.

## 8.4 Development, Diagnostics and Validation of univariate logistic regression model (ufov3 based)

Having determined the cognitive test (*ufov3*) that constitutes the best fitting model to the data, the next step was to determine the correct functional form of *ufov3* in which it should enter the model (i.e. whether the model is linear in the logit for *ufov3*), assess goodness-of-fit of the model (overall measures of fit) and perform diagnostics analysis. Before these analyses are carried out it is necessary to elaborate the concept of *covariate pattern*. Hosmer and Lemeshow (2000) recommends calculating diagnostic statistics by *covariate pattern*. All subjects/observations having the same covariate values (the same values on the predictors) constitute a particular covariate pattern. For example, in the 56 observations (relevant to drivers) there were 38 covariate patterns (i.e. there were 38 distinct/unique values of *ufov3*). The model is developed/fitted based on individual subject/observation data, however, the covariate pattern is taken into account when the fit of the model is assessed (Hosmer and Lemeshow, 2000). The covariate values (predictor values), fitted values and diagnostic statistics are the same for all subjects/observations with a particular covariate pattern, but each subject/observation has an individual outcome. The *covariate pattern* approach to assessing goodness-of-fit and diagnostics is necessary because if the number of covariate patterns is much smaller than n (the total number of observations), identification of influential and/or poorly fit covariate patterns may be missed (Hosmer and Lemeshow, 2000).

To determine the correct functional form of *ufov3* in which it should enter the model, the *scatterplot smooth* approach was used (Hosmer and Lemeshow, 2000). In this approach, first a Lowess curve is fitted to a graph between observed response (i.e. for “poor-drivers” y=1 and for “not-poor-drivers” y=0) on the y-axis and *ufov3* on the x-axis. It may be highlighted that the Lowess curve follows the general trend in the data. Ordinates corresponding to each *ufov3* value are read from this graph which are in fact probabilities of being a “poor-driver” and these probabilities are converted into logits using Equation 8.16, which are called smoothed logits. Finally a graph is drawn between the smoothed logits and ufov3 which is called a scatterplot smooth, shown in Figure 8.2.

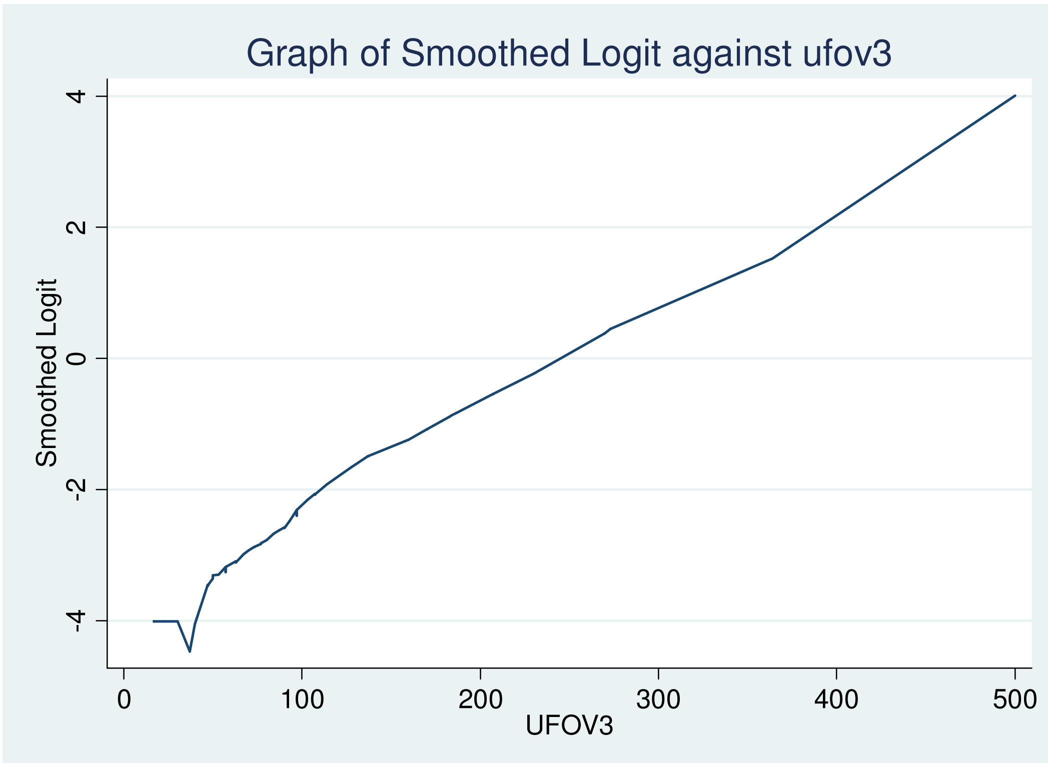



**Figure 8.2** Graph of Smoothed Logit against ufov3.

Figure 8.2 supports treating *ufov3* as linear in the logit (i.e., no transformation of *ufov3*). The results of logistic regression for the predictor ufov3 are shown in Tables 8.3 to 8.5.

**Table 8.3** Goodness-of-fit statistics for univariate logistic regression using *ufov3*

| Logistic Regression Model with *ufov3* as predictor | |
|---|---|
| Number of observations | 56 |
| Log Likelihood of Null Model | -22.966514 |
| Log Likelihood of Full Model | -14.839488 |
| Likelihood ratio Chi-square (1) | 16.25 |

| Logistic Regression Model with *ufov3* as predictor | |
|---|---|
| Prob > Chi-square | 0.0001 |
| Pseudo-$R^2$ | 0.3539 |
| AIC | 33.67898 |
| BIC | 37.72968 |
| Deviance | 29.679 |

**Table 8.4** Coefficients for univariate logistic regression using *ufov3*

| Predictor | Coefficient | Standard Error | z | P>\| z \| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| *ufov3* | 0.019135 | 0.0064765 | 2.95 | 0.003 | 0.0064412 | 0.0318288 |
| constant | -4.591905 | 1.161746 | -3.95 | 0 | -6.868885 | -2.314925 |

**Table 8.5** Odd Ratio for univariate logistic regression using *ufov3*

| Predictor | Odds Ratio | Standard Error | z | P>\| z \| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| *ufov3* | 1.019319 | 0.0066017 | 2.95 | 0.003 | 1.006462 | 1.032341 |

To see whether the logistic regression model fits the data well overall, the Pearson goodness-of-fit test was carried out, which uses the covariate patterns in the data as groups. The test is based on the idea of comparing an observed number of individuals to the number expected calculated from the fitted model. These observed (O) and expected (E) numbers are then combined to form a goodness-of-fit $\chi^2$ statistic. Large values of the statistic indicate a poor fit (with its concomitant small p-values). The Pearson goodness-of-fit test gave a chi-square (36 degrees of freedom) value of 40.71 translating to a p-value of 0.2707, which indicated that overall the model fitted the data well. Another way of graphically assessing the fit of the model is to plot a Lowess graph comparing predicted probabilities to a moving average of the proportion of cases that are 1 (i.e., “poor-drivers”). In other words a Lowess (smoother curve) graph between the observed outcome i.e. coded 1 for “poor-driver” and 0 for “not-poor-driver” on the y-axis and predicted probability of being a “poor-driver” on the x-axis. Such a graph is shown at Appendix-G, where a diagonal straight line (having equal x-axis and y-axis values) has also been embedded for reference. The graph shows the Lowess (smoother curve) curve closely following the diagonal line, indicating that the fraction of observed cases are about equal to the predicted probabilities, which manifests a good fit of the model.

In diagnostics, the influence of individual patterns is considered i.e. the individual components of the summary statistics are graphically examined. The *Delta Chi-square* statistic and the *Delta-D* (i.e Delta-Deviance) statistic allow the identification of those covariate patterns that are poorly fit (large values of *Delta Chi-square* and/or *Delta-D*). Large values of *Pregibon's dbeta* allow the identification of those covariate patterns that have a great deal of influence on the values of the estimated parameters of the model. The *Delta Chi-square* is defined as:

$$\Delta X_j^2 = X^2 - X_{(j)}^2 \tag{8.11}$$

Where $\Delta X_j^2$ is the change in the Pearson chi-square goodness of fit statistic as a result of deleting covariate pattern *j*, $X^2$ is the Pearson chi-square goodness of fit statistic computed from the full data set and $X_{(j)}^2$ is Pearson chi-square goodness of fit statistic computed when covarite pattern *j* is deleted. The *Delta-D* statistic is defined as:

$$\Delta D_j = DEV - DEV_{(j)} \tag{8.12}$$

Where $\Delta D_j$ the change in the Deviance statistic as a result of deleting covariate pattern *j*, $DEV$ is the Deviance statistic computed from the full data set and $DEV_{(j)}$ is the Deviance statistic computed when covarite pattern *j* is deleted. Thus the *Delta Chi-square* and *Delta-D* statistic provide measures of the influence of the jth covariate pattern on these summary statistics. In order to determine n (i.e. the number of observations) *Delta Chi-square* and *Delta-D* statistics, n models have to be fitted in each case which is computation intensive and time consuming. Therefore, for faster computation, linear approximations are used (Hosmer and Lemeshow, 2000). The *Pregibon's dbeta* ($\Delta\beta_j$) statistic is used to assess the influence that individual covariate patterns have on the estimated model parameters and is the difference between the estimated parameters from the full data set and the estimated parameters when a specific covariate pattern (i.e. pattern *j*) is excluded from the analysis. *Pregibon's dbeta* ($\Delta\beta_j$) represents the influence of covariate pattern *j* on all the estimated $\beta$'s simultaneously. To save computation effort and time (and avoid fitting the model n times), approximations of the measure have been developed (Hosmer and Lemeshow, 2000). The diagnostic statistics are usually plotted against predicted probabilities of the outcome of interest (i.e. predicted probability of being a "poor-driver") obtained from the model. In logistic regression, since the distribution of the diagnostic statistics under the hypothesis that the models fits is not known with certainty, therefore, reliance on visual assessment of the plots is the mainstay (Hosmer and Lemeshow, 2000). In assessing diagnostic statistics, the relative magnitude of the statistic of an observation is compared to others i.e. observations that are farther away from most of the data points warrant attention. However, values of *Delta Chi-square* and *Delta-D* statistic that not much greater than 4 are considered not high and the *Pregibon's dbeta* ($\Delta\beta_j$) should be larger than 1.0 for an individual covariate pattern to have an effect on the estimated coefficients (Hosmer and Lemeshow, 2000).

Since the summary statistic (Pearson goodness-of-fit test) indicated that overall the model fitted well, therefore it was not expected of diagnostics to find a large number of covariate patterns being poorly fit. The graph between *Delta Chi-square* statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") and between *Delta-D* statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") in Appendix-G, shows driver "O13" as some distance from the balance of plotted-data and has values of both these statistics much greater than 4 (i.e. partly based on visual impression from the graphs and partly based on numeric values of these two statistics). Clearly driver "O13" is poorly fit by the model, as she was classified as "poor-driver" whereas her score on ufov3 was favourable (i.e. 67). Aside from the observation of driver "O13", the plots show that the model fits reasonably well.

The graph between *Pregibon's dbeta* ($\Delta\beta_j$) statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") in Appendix-G, shows driver "O13" and "O14" to be some distance from the balance of plotted-data, although both had values of *Pregibon's dbeta* ($\Delta\beta_j$) less than 1.0, still it was decided to investigate these points further. This information can also be interpreted from the graph between *Delta Chi-square* ($\Delta X_j^2$) statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") (see Appendix-G), where the size of the symbols is proportional to *Pregibon's dbeta* ($\Delta\beta_j$) statistic. This graph shows that not all influential points (large *Pregibon's dbeta*) are outliers (large *Delta Chi-square* or *Delta-D*).

**Table 8.6** Estimated coefficients from all data, the percent change when the covariate pattern is deleted, values of Goodness-of-Fit statistic, and Odds ratio for each Model.

| Variable | All Data | Deleting driver "O13" | Deleting driver "O14" | Deleting driver "O13"&"O14" |
|---|---|---|---|---|
| *ufov3* | 0.019135 | 29 % | 21 % | 68 % |
| constant | -4.59190 | -26 % | -11 % | -50% |
| Goodness-of-fit | | | | |
| Deviance | 29.679 | 22.266 | 27.160 | 19.076 |
| Chi-square goodness of fit statistic | 40.71 | 16.69 | 44.52 | 15.35 |
| Odds Ratio (based on a unit change in *ufov3*) | | | | |
| *ufov3* | 1.019319 | 1.02498 | 1.023524 | 1.032687 |

**Table 8.7** Predicted probability of being a "poor-driver" and Observed outcome along with ufov3 scores for the two influential observations.

| Driver Code | Age | *ufov3* | Observed outcome (1="poor-driver", 0 ="not-poor-driver") | Predicted probability of being "poor-driver" |
|---|---|---|---|---|
| "O13" | 64.2 | 67 | 1 | 0.0352345 |
| "O14" | 76.9 | 273 | 0 | 0.652932 |

A down side of the diagnostic statistic *Pregibon's dbeta* ($\Delta\beta_j$) is that it is a summay measure of change that occurs in all coefficients in the model simultaneously (and is a linear approximation as well), therefore it was necessary to rerun the model after deleting certain covariate patterns to gauge the exact change in the coefficients. Therefore, after identification of drivers "O13" and "O14" as influential, the model was re-run with observation of driver "O13" and "O14" individually deleted one at a time and then simultaneously deleted and the actual impact on the coefficients assessed along with calculation of goodness-of-fit statistics (i.e. Deviance and Chi-square goodness of fit statistic) shown in Table 8.6. By deleting observation of driver “O13” and “O14” one at a time, the coefficient of *ufov3* increases by 29 and 21 percent respectively compared to the model fitted to all data (the combined effect is even greater i.e. increase of 68 percent in the value of the coefficient and substantial decrease in the fit measures Deviance and Chi-square goodness of fit statistic). If we examine the *Observed outcome* and *Predicted probability of being “poor-driver”* of these drivers from Table 8.7, it becomes clear why these individuals have a relatively large impact on the estimated coefficients. In general the model predicts that the odds of being a “poor-driver” increase with increasing score on *ufov3*, but driver “O13” is a “poor-driver” (from observed outcome from Table 8.7) despite her relatively low score on *ufov3* (and consequent low *Predicted probability of being “poor-driver”* from Table 8.7). On the other hand, driver “O14” is a “not-poor-driver” (from observed outcome from Table 8.7) despite his relatively high score on *ufov3* (and consequent high *Predicted probability of being “poor-driver”* from Table 8.7). Table 8.6 also corroborates the findings from the diagnostic graphs that driver “O13” is poorly fit by the model, whereas driver “O14” is not a relatively poor fit, which can be seen by comparing the Deviance and Chi-square goodness of fit statistic of the models relevant to all data and those from deleting drivers “O13” and “O14” i.e. the difference (decrease in values) between the Deviance and Chi-square goodness of fit statistic of the all-data model and the model obtained by deleting driver “O13” is substantial compared to the difference of the all-data model and the model obtained by deleting driver “O14”.

It was expected of some drivers to exhibit outcomes that were contrary to the general findings of the model, so this could not be formed as a sufficient basis for excluding these observations in fitting the model. By deleting these two observations, there was no substantial change in the conclusions and the estimated odds ratio for *ufov3* changed in the direction of making the estimated effects stronger as evident from Table 8.6. The estimated degree of association changed but our conclusions did not change substantially. Also, since no data-entry errors were found relevant to these two observations and were considered plausible as well, therefore, it was decided to keep these observations in the model.

Univariate logistic regression model is given by:

$$\pi = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}} \tag{8.13}$$

Where $\pi$ is the probability of the outcome of interest (i.e. probability of being a "poor-driver"), $\beta_0$ is the constant, and $\beta_1$ is the coefficient of the predictor *x* (i.e. *ufov3*). The logistic regression model developed using *ufov3* as a predictor was:

$$\hat{\pi} = \frac{1}{1+e^{-(-4.591905+0.019135\ ufov3)}} \tag{8.14}$$

The odds of an outcome of interest happening is defined as the probability that the outcome of interest occurs divided by the probability that outcome of interest does not occur. Therefore the odds that a driver is a "poor-driver" is:

$$Odds = \frac{\hat{\pi}}{1-\hat{\pi}} = e^{\hat{\beta}_0+\hat{\beta}_1 x} \tag{8.15}$$

The $\log_e$ of the *Odds* is called the logit and is given by:

$$\log_e\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x \tag{8.16}$$

A useful inference in logistic regression is through the use of Odds ratio. The Odds ratio is formulated to determine how the odds of the event of interest (i.e. the odds of being a "poor-driver") increases as certain changes in predictor variable values occur (for example an increase in *ufov3* of say $\eta$). The Odds ratio is the ratio of the odds of the event of interest (i.e. the odds of being a "poor-driver") at condition 2 to that of condition 1 in the predictors and is given by:

$$Odds\ ratio = \left[\frac{\hat{\pi}}{1-\hat{\pi}}\right]_2 \Big/ \left[\frac{\hat{\pi}}{1-\hat{\pi}}\right]_1 \tag{8.17}$$

Therefore, the ratio reflecting the increase in odds of being a "poor-driver" when scores on the *ufov3* increase by $\eta$ is given by (using Equations 8.15 and 8.17):

$$= \frac{e^{-4.591905+0.019135(ufov3+\eta)}}{e^{-4.591905+0.019135\ ufov3}} = e^{0.019135(\eta)} \tag{8.18}$$

When $\eta$ is 1, an Odds ratio of 1.019319 (same as in Table 8.6) is obtained (using Equation 8.18) and a $\eta$ of 40 gives an Odds ratio of 2.15. An increase in *ufov3* of 1 (i.e.

$\eta$ = 1) is not meaningful keeping in view the range of variation of *ufov3*. This means that the odds of being a "poor-driver" increase over 2.15 fold for drivers having an *increased* (i.e. difference on the higher side) *ufov3* score of 40 compared with the *ufov3* scores of other drivers or in other words, the odds of being a "poor-driver" increase 115 percent for drivers having an *increased* (i.e. difference on the higher side) *ufov3* score of 40 compared with the *ufov3* scores of other drivers.

The logistic regression model of Equation 8.14, which is based on the cognitive test *ufov3* gives an estimate of $\hat{\pi}$, which is the estimated probability of the outcome of interest (i.e. probability of being a "poor-driver") and is a continuous measure. To convert this continuous measure into a dichotomous scale (i.e. "poor-driver" and "not-poor-driver"), a cut-off point on the continuous probability scale $\hat{\pi}$, has to be determined as a prediction rule. If the value of the probability is greater than or equal to this cutoff-point, the prediction is that of a "poor-driver" and if it is less, it translates to a "not-poor-driver". As driver status is being predicted/judged through the *ufov3* test, in using a particular cut-off value of the probability, four kinds of results are possible as shown in Figure 8.3. Two of these results are correct (true) and two wrong (false). From Figure 8.3, the test (cognitive test i.e. *ufov3*) gives correct results when it is positive in the "poor-driver" status (True positive) or negative in the "not-poor-driver" status (true negative). On the other hand, the test is misleading if it is positive in the "not-poor-driver" status (false positive) or negative in the "poor-driver" status (false negative). Certain characteristics are defined relevant to the diagnostic test (cognitive test *ufov3*), these are (with reference to Figure 8.3):

$$Sensitivity = \frac{a}{a+c} \tag{8.19}$$

$$Specificity = \frac{d}{b+d} \tag{8.20}$$

$$Positive\ predictive\ value\,(+PV) = \frac{a}{a+b} \tag{8.21}$$

$$Negative\ predictive\ value\,(\text{-}PV) = \frac{d}{c+d} \tag{8.22}$$

$$1 - Specificity = 1 - \frac{d}{b+d} = \frac{b}{b+d} \quad (8.23)$$

| | | Driver Status | |
|---|---|---|---|
| | | "poor-driver" (positive) | "not-poor-driver" (negative) |
| Test (cognitive test i.e. *ufov3*) | Value of predicted Probability ≥ cut-off (the test is positive) | True positive (a) | False positive (b) |
| | Value of predicted Probability < cut-off (the test is negative) | False negative (c) | True negative (d) |

**Figure 8.3** The relationship between a diagnostic test (cognitive test i.e. *ufov3*) result and driver status (a, b, c and d are the number of drivers falling under the relevant category).

Sensitivity is the fraction of "poor-drivers" that are correctly classified by the test (i.e. positive on the cognitive test *ufov3*). Sensitivity is also known as the True-positive rate. Specificity is the fraction of "not-poor-drivers" that are correctly classified by the test (i.e. negative on the cognitive test *ufov3*). The quantity (1-Specificity) given by Equation 8.23 is known as the False-positive rate. Positive predictive value is the probability of being a "poor-driver" with a positive test result and negative predictive value is the probability of being a "not-poor-driver" with a negative test result. In other words, for example, the positive predictive value answers the question, "If a driver's test (cognitive test *ufov3*) result is positive, what are the chances that the driver will be a "poor driver?" It is desirable to have a test (cognitive test i.e. *ufov3*) that is highly sensitive and highly specific, but usually this is not possible. There is a trade-off between sensitivity and specificity of a diagnostic test (cognitive test i.e. *ufov3*), and both are affected by the decision of where the cut-off point is placed on the continuum of predicted probabilities. Consequently, sensitivity can be increased only at the expense of specificity (or vice versa).

To illustrate, as an extreme case, if the cut-point (on the predicted probability continuum) is set too low, sensitivity of the test will be very high resulting in the classification of almost everyone as a “poor-driver” with a lot of “not-poor-drivers” being erroneously classified as “poor drivers” thus resulting in low specificity for the test (i.e. cognitive test *ufov3*). Whereas if the cut-point is set too high, specificity of the test will be very high resulting in the classification of almost everyone as a “not-poor-driver” with a lot of “poor-drivers” being erroneously classified as “not-poor drivers”, thus resulting in low sensitivity for the test. In both these extreme cases, the predictive tool has no value at all. The goal should be to choose a cut-point so as to produce the fewest errors of either type. Often, there is no ideal cut-pint that will result in perfect classification. Therefore, a conscious effort should be made to minimize error of one type rather than the other.

Table in Appendix-G shows the Sensitivity, Specificity and (1- Specificity) for different cut-points (predicted probabilities used as cut-points). To classify a test result as positive (or negative), Sensitivity and Specificity rely on a single cut-point. The area under ROC (Receiver Operating Characteristic) curve gives a more complete description of classification accuracy. It plots the True-positive rate (Sensitivity) against False-positive rate (1-Specificity) for an entire range of different cut-points (Figure 8.4). The area under ROC curve varies from 0 to 1 and is a measure of the test’s ability to discriminate between “poor-drivers” and “not-poor-drivers”. An area under ROC curve of 0.5 or less means that the predictions are no better than random guessing (Kutner et al., 2005; Streiner & Norman, 2003). Well discriminating tests crowd towards the upper left corner of the ROC curve (obviously increasing the area under curve) because as sensitivity progressively increases (by lowering the cut-point), there is a small or no loss in specificity until the sensitivity reaches a very high level.

The best cut-point is generally at the shoulder of the ROC curve, unless there are specific reasons for minimizing either false negatives or false positives (Fletcher and Fletcher, 2005). According to Streiner & Norman (2003), the cut-point at the shoulder of the ROC results in the smallest overall error rate, but under certain circumstances, it may be preferable to move this point (shoulder point), either up or down , even though the number of false positive cases may increase faster than the number of true positive cases.

An area under RC curve of 0.8659 was obtained (Figure 8.4) which is considered excellent discrimination (Hosmer and Lemeshow, 2000).

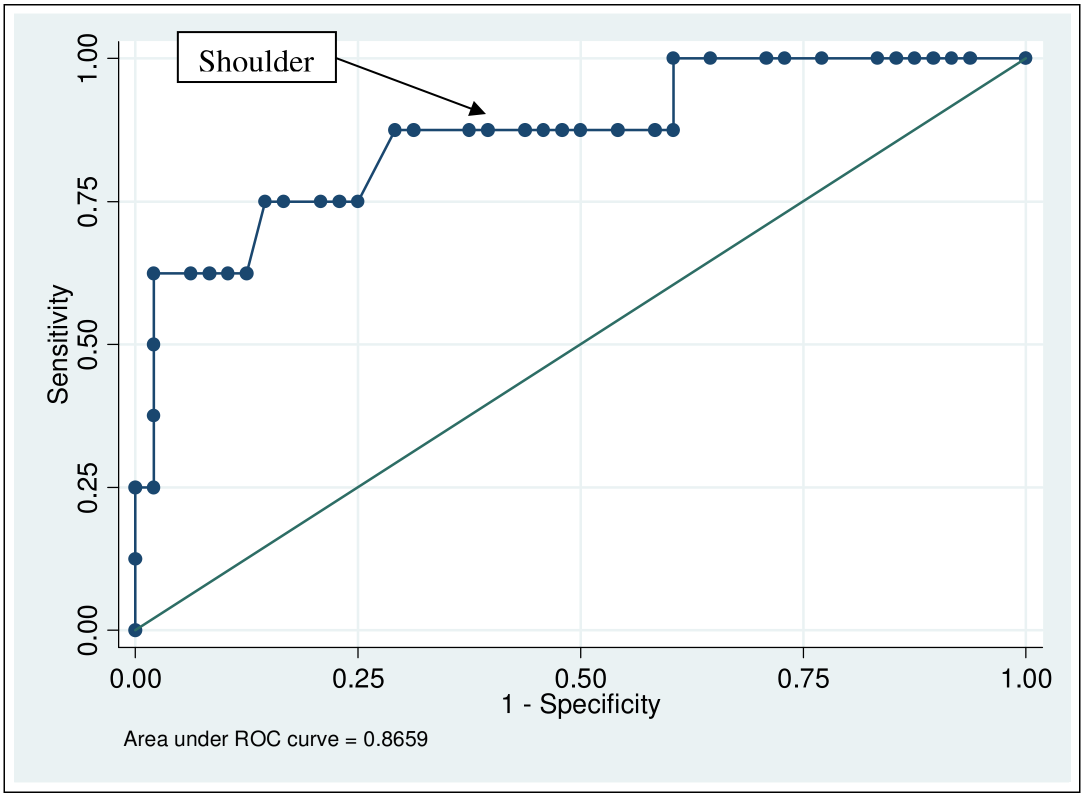



**Figure 8.4** Receiver Operating Characteristic (ROC) curve for different cut-pints (predicted probabilities used as cut-points). Area under ROC curve is 0.8659.

Incrementally, different values of predicted proabability (of being a "poor driver") that served as cut-off points were used. If the predicted probability was greater than or equal to a particular cut-off point, the driver was classified as a "poor-driver" by the model and if less he/she was classified as a "not-poor-driver". This classification from the model was compared with the actual status of the driver (whether a "poor-driver" or a "not-poor-driver") and the correct classification rate determined for that particular cut-point. The value of the cut-point that resulted in the highest correct classification rate was selected as the appropriate cut-off point. In the logistic regression model based on ufov3, a cut-point of 0.4 provided the highest correct classification rate of drivers.

In Appendix-G, the table relevant to Sensitivity, Specificity and (1-Specificity) for different cut-points (predicted probabilities used as cut points), graph between Sensitivity/Specificity versus cut-points and the ROC curve can also guide in the selection of an appropriate cut-point. Figure 8.5 shows the driver classification obtained as a result of using a cut-point value of 0.4.

| | | Driver Status | | |
|---|---|---|---|---|
| | | "poor-driver" (positive) | "not-poor-driver" (negative) | Total |
| Test (cognitive test i.e. *ufov3*) | Value of predicted Probability $\geq$ 0.4 (cut-point) (the test is positive) | 5 | 1 | 6 |
| | Value of predicted Probability < 0.4 (cut-point) (the test is negative) | 3 | 47 | 50 |
| Total | | 8 | 48 | 56 |

**Figure 8.5** The relationship between cognitive test (i.e. *ufov3*) result and driver status classification showing the number of drivers in each category using a cut-point value of predicted probability of 0.4.

Other parameters relevant to the table are calculated as follows:

$$Sensitivity = \frac{5}{5+3} \times 100 = 62.5\% \tag{8.24}$$

$$Specificity = \frac{47}{1+47} \times 100 = 97.92\% \tag{8.25}$$

$$Positive\ predictive\ value\,(+PV) = \frac{5}{5+1} \times 100 = 83.33\% \tag{8.26}$$

$$\textit{Negative predictive value (-PV)} = \frac{47}{3+47} \times 100 = 94\% \tag{8.27}$$

$$\textit{Correct Classification} = \frac{5+47}{56} \times 100 = 92.86\% \tag{8.28}$$

92.89 percent of the drivers were correctly classified as either being “poor-drivers” or “not-poor-drivers” by the diagnostic test (cognitive test i.e. *ufov3*). Only one “not-poor-driver” out of 48 was classified as a “poor-driver” by the test and three “poor-drivers” out of 8 as “not-poor-drivers”. In total, out of the 56 drivers only 4 drivers were misclassified by the test (cognitive test i.e. *ufov3*). Positive and negative predictive values of the test were high.

Validation of a model is crucial especially if the model is to be employed to predict the outcome (whether a “poor-driver” or “not-poor-driver”) of future subjects (drivers). The aim of validation procedures in logistic regression is typically to assess discrimination performance, which is the ability of a test to bifurcate a population into those who will experience the outcome of interest (i.e. “poor-drivers”) and those who will not (“not-poor-drivers”). A review of validation techniques has been provided in Section 7.4.7. Owing to the small sample size and as suggested by Harrell et al. (1996), the *K-fold cross-validation* was adopted. In the *K-fold cross-validation* technique, the data are split into $K$ roughly equal parts. Out of the $k$ = 1, 2, ….., $K$ parts, the $k^{th}$ part is used as the *validation set* and the remaining $K$-1 parts are used as the *model-building set*. When the data split is such that $K$= n, it becomes the PRESS statistic variant of data-splitting (Montgomery et al., 2006; Weisberg, 2005; Snee, 1977) in multiple linear regression, explained in Section 7.4.7. In logistic regression, $K$= J , where J ( $J \leq n$) is the number of covariate patterns present in the data, because predicted values (fitted values) are the same for the same covariate pattern. Therefore, almost all the data (except one covariate pattern) is used to estimate the coefficients of the model. A model is estimated from all covariate patterns except covariate pattern $j$. Then this model is used to predict the fitted values (predicted probability) $\hat{\pi}_{(j)}$ of covariate pattern $j$. Thus a model is used to predict the fitted values of a covariate pattern whose data was not used in the development of the model. This technique is also called LOO (leave-one-out) (Bautista et al., 1999). The use of the covariate patterns in validation has been advocated by Houwelingen and Cessie

(1990) and of the technique in general by Steyerberg et al. (2001) and Altman and Royston (2000). Unlike multiple linear regression, there are no computationally-economic formulas available and so the model has to be fitted a J number of times (i.e. equal to the number of covariate patterns). In model validation, the discrimination ability obtained from a model using the training sample usually decreases in the validation sample because the discrimination estimates (of the training sample) are derived from the same data that was used to fit the model (Efron, 1986). The LOO (leave-one-out) technique was used jointly with the area under ROC (Receiver Operating Characteristic) curve for validating the logistic regression model (Bautista et al., 1999), because typically ROC curves are used to evaluate model discrimination. First, the area under ROC was calculated from the model based on all data and then it was compared with the area under ROC curve obtained from the predicted probabilities using the LOO technique (leave-one-out). The number of covariate patterns in the data was 38 (because there were 38 unique values of *ufov3* among the 56 drivers), therefore for the LOO technique, the model was fitted 38 times, each time leaving a particular covariate pattern out. The area under ROC curve obtained using all the data was 0.8659 (see Figure 8.4) and the area under ROC curve obtained using the LOO technique was 0.7904 (ROC curve and predicted probabilities obtained using LOO are at Appendix-G).

The statistical software Stata® was used to test the equality of both areas using a procedure suggested by DeLong et al. (DeLong et al., 1988 cited in Stata® 10 manual), in which the null hypothesis is $H_0$: both areas are equal. Although the area 0.8659 is larger than 0.7904, the chi-squared test yielded a p-value of 0.0785, suggesting that there is no significant difference between the two areas, showing that the discriminatory ability of the logistic regression model did not decrease significantly using the LOO technique and therefore there was reasonably strong evidence that the model would be a satisfactory predictor for determining the status (whether “poor-driver” or “not-poor-driver”) of future drivers, based on scores of *ufov3* test.

## 8.5 Development, Diagnostics and Validation of Univariate Logistic Model (Age Based)

In order to determine the extent to which age alone can discriminate between "poor-drivers" and "not-poor-drivers", it was decided to use age as a predictor in a logistic regression model. To develop the model, it was necessary to determine the correct functional form of *age* in which it should enter the model (i.e. whether the model is linear in the logit for *age*), assess goodness-of-fit of the model (overall measures of fit) and perform diagnostics analysis. There were 52 covariate patterns (i.e. there were 52 distinct/unique values of *age*) in this model. To determine the correct functional form of *age* in which it should enter the model, the *scatterplot smooth* approach was used (Hosmer and Lemeshow, 2000). The scatterplot smooth, where the results are plotted on the logit scale is shown in Appendix-G. The graph essentially supports linearity of the logit in *age* (i.e., no transformation of *age*). The results of logistic regression for the predictor age are shown in Tables 8.8 to 8.10.

**Table 8.8** Goodness-of-fit statistics for univariate logistic regression using *age.*

| Logistic Regression Model with age as predictor | |
|---|---|
| Number of observations | 56 |
| Log Likelihood of Null Model | -22.966514 |
| Log Likelihood of Full Model | -13.385772 |
| Likelihood ratio Chi-square (1) | 19.16 |
| Prob > Chi-square | 0.0000 |
| Pseudo-$R^2$ | 0.4172 |
| AIC | 30.77154 |
| BIC | 34.82225 |
| Deviance | 26.772 |

**Table 8.9** Coefficients for univariate logistic regression using *age*.

| Predictor | Coefficient | Standard Error | z | P>\| z \| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| *age* | 0.1654765 | 0.0659499 | 2.51 | 0.012 | 0.0362171 | 0.2947359 |
| constant | -12.49311 | 4.666546 | -2.68 | 0.007 | -21.63937 | -3.346847 |

**Table 8.10** Odd Ratio for univariate logistic regression using *age*

| Predictor | Odds Ratio | Standard Error | z | P>\| z \| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| *age* | 1.179955 | 0.0778179 | 2.51 | 0.012 | 1.036881 | 1.342772 |

The overall fit of the model to see whether the logistic regression model overall fits the data well was carried out using the Hosmer and Lemeshow goodness-of-fit test. In this test, the probability of being a "poor-driver" is calculated from the model for each observation and the resulting values are arranged in increasing order. The range of probability values is then divided into subgroups (for example, deciles). For each sub-group the observed number (O) of "poor-drivers" and "not-poor-driver" is noted and the corresponding expected (E) number of "poor-drivers" are calculated by adding the probabilities (of being a "poor-driver") for each individual in each subgroup and the expected number of "not-poor-drivers" is the complement of this summation (i.e., 1-(sum of probabilities)). The goodness-of-fit statistic is calculated from the Pearson Chi-square statistic as:

$$\text{Goodness-of-fit } \chi^2 = \sum \frac{(O-E)^2}{E} \tag{8.29}$$

The degrees of freedom for this statistic is the number of sub-groups minus two. Large values of the statistic indicate a poor fit (with its concomitant small p-values). The Hosmer and Lemeshow goodness-of-fit test gave a chi-square (5 degrees of freedom) value of 2.15 translating to a p-value of 0.8285, which indicated that overall the model fitted the data well. Another way of graphically assessing the fit of the model is to plot a Lowess graph comparing predicted probabilities to a moving average of the proportion of cases that are 1 (i.e., "poor-drivers"). In other words a Lowess (smoother curve) graph between the observed outcome i.e. coded 1 for "poor-driver" and 0 for "not-poor-driver" on the y-axis and predicted probability of being a "poor-driver" on the x-axis. Such a graph is shown at Appendix-G, where a diagonal straight line (having equal x-axis and y-axis values) has also been embedded for reference. The graph shows the Lowess (smoother curve) curve following the diagonal line quite in proximity, indicating that the fraction of observed cases are about equal to the predicted probabilities, which manifests a good fit of the model.

In diagnostics, the influence of individual patterns is considered i.e. the individual components of the summary statistics are graphically examined. Since the summary statistic (Hosmer and Lemeshow goodness-of-fit test) indicated that overall the model fitted well, therefore it was not expected of diagnostics to find a large number of covariate patterns being poorly fit. The graph between *Delta Chi-square* statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") and between *Delta-D* statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") in Appendix-G, shows driver "O13" and "O44" some distance from the balance of plotted-data and have values of both these statistics slightly greater than 4 (i.e. partly based on visual impression from the graphs and partly based on numeric values of these two statistics). Driver "O13" and "O44" are poorly fit by the model, as they were "poor-drivers" whereas their ages were relatively not too much (i.e. 64.2 and 64.7 years respectively). Aside from the observations of driver "O13" and "O44", the plots show that the model fits reasonably well.

The graph between *Pregibon's dbeta* ($\Delta\beta_j$) statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") in Appendix-G, shows driver "O13", "O44" and "O21" to be some distance from the balance of plotted-data, although all three had values of *Pregibon's dbeta* ($\Delta\beta_j$) less than 1.0, still it was decided to investigate these points further. This information can also be interpreted from the graph between *Delta Chi-square* ($\Delta X_j^2$) statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") (see Appendix-G), where the size of the symbols is proportional to *Pregibon's dbeta* ($\Delta\beta_j$) statistic. This graph shows that not all influential points (large *Pregibon's dbeta*) are outliers (large *Delta Chi-square* or *Delta-D*).

**Table 8.11** Estimated coefficients from all data, the percent change when the covariate pattern is deleted, values of Goodness-of-Fit statistic, and Odds ratio for each Model.

| Variable | All Data | Deleting driver "O13" | Deleting driver "O44" | Deleting driver "O21" | Deleting drivers "O13","O44" & "O21" |
|---|---|---|---|---|---|
| *age* | 0.1654765 | +19 % | +17 % | +21 % | + 96 % |
| constant | -12.49311 | -20 % | - 18 % | -18 % | -93 % |
| Goodness-of-fit | | | | | |
| Deviance | 26.772 | 22.388 | 22.565 | 24.319 | 14.333 |
| Chi-square goodness of fit statistic (Hosmer & Lemeshow) | 2.15 | 0.65 | 0.62 | 2.29 | 1.69 |
| Odds Ratio (based on a unit change in age i.e. 1 year) | | | | | |
| *age* | 1.179955 | 1.217456 | 1.213627 | 1.222437 | 1.383792 |

**Table 8.12** Predicted probability of being a "poor-driver" and Observed outcome along with age for the three influential observations.

| Driver Code | Age | Observed outcome (1="poor-driver", 0 ="not-poor-driver") | Predicted probability of being "poor-driver" |
|---|---|---|---|
| "O13" | 64.2 | 1 | 0.1335977 |
| "O44" | 64.7 | 1 | 0.1434682 |
| "O21" | 79.4 | 0 | 0.656047 |

A down side of the diagnostic statistic *Pregibon's dbeta* ($\Delta\beta_j$) is that it is a summary measure of change that occurs in all coefficients in the model simultaneously (and is a linear approximation as well), therefore it was necessary to rerun the model after deleting certain covariate patterns to gauge the exact change in the coefficients. Therefore, after identification of drivers "O13", "O44" and "O21" as influential, the model was re-run with observation of driver "O13", "O44" and "O21" individually deleted one at a time and then simultaneously deleted and the actual impact on the coefficients assessed along with calculation of goodness-of-fit statistics (i.e. Deviance and Hosmer and Lemeshow chi-square goodness of fit statistic) shown in Table 8.11. By deleting observation of driver "O13", "O44" and "O21" one at a time, the coefficient of *age* increases by 19, 17 and 21 percent respectively compared to the model fitted to all data (the combined effect is even greater i.e. increase of 96 percent in the value of the coefficient and substantial decrease in the fit measures Deviance and Hosmer and Lemeshow chi-square goodness of fit statistic). If we examine the *Observed outcome* and *Predicted probability of being "poor-driver"* of these drivers from Table 8.12, it becomes clear why these individuals have a relatively large impact on the estimated coefficients. In general the model predicts that the odds of being a "poor-driver" increase with increasing *age*, but driver "O13" and "O44" are "poor-drivers" (from observed outcome from Table 8.12) despite their relatively low age (and consequent low *Predicted probability of being "poor-driver"* from Table 8.12). On the other hand, driver "O21" is a "not-poor-driver" (from observed outcome from Table 8.12) despite her relatively high age (and consequent high *Predicted probability of being "poor-driver"* from Table 8.12). Table 8.11 also corroborates the findings from the diagnostic graphs that driver "O13" and "O44" are poorly fit by the model, whereas driver "O21" is not a relatively poor fit, which can be seen by comparing the Deviance and Hosmer and Lemeshow chi-square goodness of fit statistic of the models relevant to all data and those from deleting drivers "O13", "O44" and "O21" i.e. the difference (decrease in values) between the Deviance and Hosmer and Lemeshow chi-square goodness of fit statistic of the all-data model and the model obtained by deleting driver "O13" and "O44" is substantial compared to the difference of the all-data model and the model obtained by deleting driver "O21".

It was expected of some drivers to exhibit outcomes that were contrary to the general findings of the model, so this could not be formed as a sufficient basis for excluding these observations in fitting the model. By deleting these three observations, there was no substantial change in the conclusions and the estimated odds ratio for *age* changed in the direction of making the estimated effects stronger as evident from Table 8.11. The estimated degree of association changed but our conclusions did not change substantially. Also, since no data-entry errors were found relevant to these two observations and were considered plausible as well, therefore, it was decided to keep these observations in the model.

The ratio reflecting the increase in odds of being a "poor-driver" when *age* increases by $\eta$ is given by (using Equations 8.15 and 8.17):

$$= \frac{e^{-12.49311+0.1654765(age+\eta)}}{e^{-12.49311+0.1654765(age)}} = e^{0.1654765(\eta)} \tag{8.30}$$

When $\eta$ is 1, an Odds ratio of 1.179955 (same as in Table 8.11) is obtained (using Equation 8.30) and a $\eta$ of 10 gives an Odds ratio of 5.23. An increase in *age* of 1 year (i.e. $\eta = 1$) is not meaningful keeping in view the range of variation of age. This means that the odds of being a "poor-driver" increase over 5.23 fold for drivers having an *increased* (i.e. difference on the higher side) age of 10 years compared to the *age* of other drivers or in other words, the odds of being a "poor-driver" increase 423 percent for drivers having an *increased* (i.e. difference on the higher side) *age* of 10 compared to the age of other drivers.

An area under ROC curve of 0.9062 was obtained which is considered excellent discrimination (Hosmer and Lemeshow, 2000). A cut-point of 0.56 was selected (as explained in Section 8.4) as it gave the highest overall correct classification of drivers. Figure 8.6 shows the driver classification obtained as a result of using a cut-point value of 0.56.

Using cut-pint of 0.56 on the predicted probability scale, a sensitivity of 50 percent, a specificity of 97.92 percent, a positive predictive value of 80 percent, a negative predictive value of 92.16 percent and a correct classification rate of 91.07 percent was obtained. That is, 91.07 percent of the drivers were correctly classified as either being "poor-drivers" or "not-poor-drivers" by the diagnostic test (*age*). Only one "not-poor-driver" out of 48 was classified as a "poor-driver" by the test and four "poor-drivers" out of 8 as "not-poor-drivers". In total, out of the 56 drivers only 5 drivers were misclassified by the test (i.e. *age*). Positive and negative predictive values of the test were high.

For validating the model, the methodology of Section 8.4 was used. The LOO (leave-one-out) technique was used jointly with the area under ROC (Receiver Operating Characteristic) curve for validating the logistic regression model (Bautista et al., 1999), because typically ROC curves are used to evaluate model discrimination.

| | | Driver Status | | |
|---|---|---|---|---|
| | | "poor-driver" (positive) | "not-poor-driver" (negative) | Total |
| Test ( i.e. *age*) | Value of predicted Probability $\geq 0.56$ (cut-point) (the test is positive) | 4 | 1 | 5 |
| | Value of predicted Probability $< 0.56$ (cut-point) (the test is negative) | 4 | 47 | 51 |
| Total | | 8 | 48 | 56 |

**Figure 8.6** The relationship between Test (i.e. *age*) and driver status classification showing the number of drivers in each category using a cut-point value of predicted probability of 0.56.

First, the area under ROC was calculated from the model based on all data and then it was compared with the area under ROC curve obtained from the predicted probabilities using the LOO technique (leave-one-out). The number of covariate patterns in the data were 52, therefore for the LOO technique, the model was fitted 52 times, each time leaving a particular covariate pattern out. The area under ROC curve obtained using all the data was 0.9062 and the area under ROC curve obtained using the LOO technique was 0.8724 (ROC curve and predicted probabilities obtained using LOO are at Appendix-G). The statistical software Stata® was used to test the equality of both areas using a procedure suggested by DeLong et al. (DeLong et al., 1988 cited in Stata® 10 manual), in which the null hypothesis is $H_0$: both areas are equal. Although the area 0.9062 is larger than 0.8724, the chi-squared test yielded a p-value of 0.1026, suggesting that there is no significant difference between the two areas, showing that the discriminatory ability of the logistic regression model did not decrease significantly using the LOO technique and therefore there was reasonably strong evidence that the model would be a satisfactory predictor for determining the status (whether "poor-driver" or "not-poor-driver") of future drivers, based on *age*.

## 8.6 Development, Diagnostics and Validation of Univariate Logistic Model (based on Composite Cognitive Measure)

To explore if a cognitive measure that is based on a composite score of all nine neuropsychological tests was a better discriminator than the *ufov3* test alone, a univariate logistic regression model based on such a cognitive measure was developed. To formulate a composite score of the nine cognitive tests (*trail, clock, rey-copy, rey-recall, dichotic, paper, ufov1, ufov2, ufov3*), first they were standardized (i.e., the mean of each cognitive test was subtracted from its respective value and then divided by the sample standard deviation of the test, resulting in each test having a mean of zero and a standard deviation of 1. Values larger than the mean appeared as positive and smaller as negative. Since standardization is a linear transformation, the distribution of each cognitive test was preserved (because the proportionality of inter-score distance was preserved, therefore there was no distortion in shape of distribution). The polarity of *clock*, *rey-copy*, *rey-recall* and *paper* was changed (i.e., their standardized scores were multiplied by -1) so that greater scores on these tests also translates to unfavourable performance. The standardized scores were added to form a composite score and this composite score was converted into a T-score by means of the formula:

$$Tscore\_cognitive = \bar{\mathrm{X}}' + z\left(\mathrm{SD}'\right) \qquad (8.31)$$

Where,

$\bar{\mathrm{X}}'$ = desired mean of the variable i.e. 100.

$\mathrm{SD}'$ = desired standard deviation of the variable i.e. = 15.

$z$ = the *standardized* form of the composite score.

$Tscore\_cognitive$ = A measure of the composite cognitive score such that its mean is now 100 and standard deviation 15.

The choice of the mean (i.e. 100) and standard deviation (i.e. 15) ensured that we do not end up with negative values on $Tscore\_cognitive$. Equation 8.31 is a linear transformation and does not affect the distribution of the composite cognitive score. Higher scores on $Tscore\_cognitive$ translated to unfavourable performance (this orientation was necessary to make it compatible with the other models for comparison purposes).

In order to determine the extent to which this composite cognitive measure of nine tests (i.e. $Tscore\_cognitive$) could discriminate between "poor-drivers" and "not-poor-drivers", it was decided to use $Tscore\_cognitive$ as a predictor in a logistic regression model (using the *pessimistic* scenario). To develop the model, it was necessary to determine the correct functional form of $Tscore\_cognitive$ in which it should enter the model (i.e. whether the model is linear in the logit for $Tscore\_cognitive$), assess goodness-of-fit of the model (overall measures of fit) and perform diagnostics analysis. There were 56 covariate patterns (i.e. there were 56 distinct/unique values of $Tscore\_cognitive$) in this model. To determine the correct functional form of $Tscore\_cognitive$ in which it should enter the model, the *scatterplot smooth* approach was used (Hosmer and Lemeshow, 2000). The scatterplot smooth, where the results are plotted on the logit scale is shown in Appendix-G. The graph essentially supports linearity of the logit in $Tscore\_cognitive$ (i.e., no transformation of $Tscore\_cognitive$). The results of logistic regression for the predictor $Tscore\_cognitive$ are shown in Tables 8.13 to 8.15.

**Table 8.13** Goodness-of-fit statistics for univariate logistic regression using $Tscore\_cognitive$.

| Logistic Regression Model with $Tscore\_cognitive$ as predictor | |
|---|---|
| Number of observations | 56 |
| Log Likelihood of Null Model | -22.966514 |
| Log Likelihood of Full Model | -16.148665 |
| Likelihood ratio Chi-square (1) | 13.64 |
| Prob > Chi-square | 0.0002 |
| Pseudo-$R^2$ | 0.2969 |

| Logistic Regression Model with *Tscore_cognitive* as predictor | |
|---|---|
| AIC | 36.29733 |
| BIC | 40.34803 |
| Deviance | 32.297 |

**Table 8.14** Coefficients for univariate logistic regression using *Tscore_cognitive* .

| Predictor | Coefficient | Standard Error | z | P>\| z \| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| *Tscore_cognitive* | 0.1067349 | 0.0366649 | 2.91 | 0.004 | 0.034873 | 0.1785968 |
| constant | -12.91376 | 3.928084 | -3.29 | 0.001 | -20.61266 | -5.214853 |

**Table 8.15** Odd Ratio for univariate logistic regression using *Tscore_cognitive* .

| Predictor | Odds Ratio | Standard Error | z | P>\| z \| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| *Tscore_cognitive* | 1.112639 | 0.0407948 | 2.91 | 0.004 | 1.035488 | 1.195539 |

The overall fit of the model to see whether the logistic regression model overall fits the data well was carried out using the Hosmer and Lemeshow goodness-of-fit test. Large values of the statistic indicate a poor fit (with its concomitant small p-values). The Hosmer and Lemeshow goodness-of-fit test gave a chi-square (5 degrees of freedom) value of 5.19 translating to a p-value of 0.3928, which indicated that overall the model fitted the data well. Another way of graphically assessing the fit of the model is to plot a Lowess graph comparing predicted probabilities to a moving average of the proportion of cases that are 1 (i.e., "poor-drivers"). In other words a Lowess (smoother curve) graph between the observed outcome i.e. coded 1 for "poor-driver" and 0 for "not-poor-driver" on the y-axis and predicted probability of being a "poor-driver" on the x-axis. Such a graph is shown at Appendix-G, where a diagonal straight line (having equal x-axis and y-axis values) has also been embedded for reference. The graph shows the Lowess (smoother curve) curve following more or less in proximity, indicating that the fraction of observed cases are about equal to the predicted probabilities, which manifests a satisfactory fit of the model.

In diagnostics, the influence of individual patterns is considered i.e. the individual components of the summary statistics are graphically examined. Since the summary statistic (Hosmer and Lemeshow goodness-of-fit test) indicated that overall the model fitted well, therefore it was not expected of diagnostics to find a large number of covariate patterns being poorly fit. Considering all the data, the maximum value of $Tscore\_cognitive$ was 177.8352 and the minimum 82.66582, with lower scores considered favourable. The graph between *Delta Chi-square* statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") and between *Delta-D* statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") in Appendix-G, shows driver "O13", "O44", "O6" and "O15" some distance from the balance of plotted-data and have values of both these statistics almost greater than 4 (i.e. partly based on visual impression from the graphs and partly based on numeric values of these two statistics). Drivers "O13", "O44", "O6" and "O15" are poorly fit by the model, as they were classified as "poor-drivers" whereas their scores on $Tscore\_cognitive$ were relatively low (i.e. 106.9, 102.3, 100.5 and 99.4 respectively). Aside from the observations of drivers "O13", "O44", "O6" and "O15", the plots show that the model fits reasonably well.

The graph between *Pregibon's dbeta* ($\Delta\beta_j$) statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") in Appendix-G, shows driver "O21" to be some distance from the balance of plotted-data, although her value of *Pregibon's dbeta* ($\Delta\beta_j$) was less than 1.0, still it was decided to investigate this point further. This information can also be interpreted from the graph between *Delta Chi-square* ($\Delta X_j^2$) statistic and Pr (y) (i.e. predicted probability of being a "poor-driver") (see Appendix-G), where the size of the symbols is proportional to *Pregibon's dbeta* ($\Delta\beta_j$) statistic. This graph shows that not all influential points (large *Pregibon's dbeta*) are outliers (large *Delta Chi-square* or *Delta-D*).

A down side of the diagnostic statistic *Pregibon's dbeta* ($\Delta\beta_j$) is that it is a summary measure of change that occurs in all coefficients in the model simultaneously (and is a linear approximation as well), therefore it was necessary to rerun the model after deleting certain covariate patterns to gauge the exact change in the coefficients.

**Table 8.16** Estimated coefficients from all data, the percent change when the covariate pattern is deleted, values of Goodness-of-Fit statistic, and Odds ratio for each Model.

| Variable | All Data | Deleting "O13" | Deleting "O44" | Deleting "O6" | Deleting "O15" | Deleting "O21" | Deleting all five drivers |
|---|---|---|---|---|---|---|---|
| *Tscore_cognitive* | 0.10673 | 0 % | +6% | +8% | +10% | +27% | +215% |
| constant | -12.9137 | -1% | -7% | -9% | -10% | -22% | -216% |
| Goodness-of-fit | | | | | | | |
| Deviance | 32.297 | 28.736 | 27.857 | 27.491 | 27.241 | 29.365 | 6.362 |
| Chi-square goodness of fit statistic (Hosmer & Lemeshow) | 5.19 | 5.08 | 5.40 | 2.66 | 2.66 | 4.90 | 0.03 |
| Odds Ratio (based on a unit change in *Tscore_cognitive* ) | | | | | | | |
| *Tscore_cognitive* | 1.1126 | 1.1126 | 1.1194 | 1.1223 | 1.1243 | 1.1446 | 1.4004 |

**Table 8.17** Predicted probability of being a "poor-driver" and Observed outcome along with score on *Tscore_cognitive* for the five observations.

| Driver Code | Age | Observed outcome (1="poor-driver", 0 ="not-poor-driver") | *Tscore_cognitive* | Predicted probability of being "poor-driver" |
|---|---|---|---|---|
| "O13" | 64.2 | 1 | 106.9118 | 0.1820469 |
| "O44" | 64.7 | 1 | 102.3312 | 0.1201039 |
| "O6" | 82.9 | 1 | 100.5739 | 0.1016506 |
| "O15" | 69.9 | 1 | 99.41039 | 0.0908582 |
| "O21" | 79.4 | 0 | 128.7144 | 0.6952022 |

Therefore, after identification of drivers "O13", "O44", "O6" and "O15" as affecting fit of the model and "O21" as influential, the model was re-run with observation of driver

“O13”, “O44”, “O6” and “O15” and “O21” individually deleted one at a time and then simultaneously deleted and the actual impact on the coefficients assessed along with calculation of goodness-of-fit statistics (i.e. Deviance and Hosmer and Lemeshow chi-square goodness of fit statistic) shown in Table 8.16. By deleting observation of driver “O13”, “O44”, “O6” , “O15” and “O21” one at a time, the coefficient of *Tscore_cognitive* increases by 0, 6, 8, 10, and 27 percent respectively compared to the model fitted to all data (the combined effect is even greater i.e. increase of 215 percent in the value of the coefficient and substantial decrease in the fit measures Deviance and Hosmer and Lemeshow chi-square goodness of fit statistic). It may be noted that observations “O13”, “O44”, “O6” and “O15” do not have a substantial effects on the estimated coefficients as predicted by *Pregibon's dbeta* ($\Delta\beta_j$). If we examine the *Observed outcome* and *Predicted probability of being “poor-driver”* of these drivers from Table 8.17, it becomes clear why these individuals have a relatively large impact on the fit of the model and the estimated coefficients. In general the model predicts that the odds of being a “poor-driver” increase with increasing score on *Tscore_cognitive*, but drivers “O13”, “O44”, “O6” and “O15” are “poor-drivers” (from observed outcome from Table 8.17) despite their relatively low score on *Tscore_cognitive* (and consequent low *Predicted probability of being “poor-driver”* from Table 8.17). On the other hand, driver “O21” is a “not-poor-driver” (from observed outcome from Table 8.17) despite her relatively high score on *Tscore_cognitive* (and consequent high *Predicted probability of being “poor-driver”* from Table 8.17). Table 8.16 also corroborates the findings from the diagnostic graphs that drivers “O13”, “O44”, “O6” and “O15” are poorly fit by the model (to varying degrees), whereas driver “O21” is not a relatively poor fit, which can be seen by comparing the Deviance and Hosmer and Lemeshow chi-square goodness of fit statistic of the models relevant to all data and those from deleting drivers “O13”, “O44”, “O6” and “O15” i.e. the difference (decrease in values) between the Deviance and Hosmer and Lemeshow chi-square goodness of fit statistic of the all-data model and the model obtained by deleting drivers “O13”, “O44”, “O6” and “O15” is substantial (Drivers “O6” and “O15” are contributing more to the misfit than drivers “O13” and

“O44” ) compared to the difference of the all-data model and the model obtained by deleting driver “O21”.

It was expected of some drivers to exhibit outcomes that were contrary to the general findings of the model, so this could not be formed as a sufficient basis for excluding these observations in fitting the model. By deleting these five observations, there was no substantial change in the conclusions and the estimated odds ratio for $Tscore\_cognitive$ changed in the direction of making the estimated effects stronger for some of the deletion scenarios as evident from Table 8.16. The estimated degree of association changed to some extent but our conclusions did not change substantially. Also, since no data-entry errors were found relevant to these five observations and were considered plausible as well, therefore, it was decided to keep these observations in the model.

The ratio reflecting the increase in odds of being a “poor-driver” when $Tscore\_cognitive$ increases by $\eta$ is given by (using Equations 8.15 and 8.17):

$$= \frac{e^{-12.91376+0.1067349(Tscore\_cognitive+\eta)}}{e^{-12.91376+0.1067349(Tscore\_cognitive)}} = e^{0.1067349(\eta)} \qquad (8.32)$$

When $\eta$ is 1, an Odds ratio of 1.112639 (same as in Table 8.15) is obtained (using Equation 8.32) and a $\eta$ of 10 gives an Odds ratio of 2.91. An increase in $Tscore\_cognitive$ of 1 (i.e. $\eta = 1$) is not meaningful keeping in view the range of variation of $Tscore\_cognitive$. This means that the odds of being a “poor-driver” increase over 2.91 fold for drivers having an *increased* (i.e. difference on the higher side) score of 10 on $Tscore\_cognitive$ compared to the score of other drivers or in other words, the odds of being a “poor-driver” increase 191 percent for drivers having an *increased* (i.e. difference on the higher side) score of 10 on $Tscore\_cognitive$ compared to the score of other drivers.

An area under ROC curve of 0.9010 was obtained which is considered excellent discrimination (Hosmer and Lemeshow, 2000). A cut-point of 0.49 was selected (as explained in Section 8.4) as it gave the highest overall correct classification of drivers. Figure 8.7 shows the driver classification obtained as a result of using a cut-point value of 0.49.

| | | Driver Status | | |
|---|---|---|---|---|
| | | "poor-driver" (positive) | "not-poor-driver" (negative) | Total |
| Test ( i.e. *Tscore_cognitive* ) | Value of predicted Probability ≥ 0.49 (cut-point) (the test is positive) | 4 | 2 | 6 |
| | Value of predicted Probability < 0.49 (cut-point) (the test is negative) | 4 | 46 | 50 |
| Total | | 8 | 48 | 56 |

**Figure 8.7** The relationship between Test (i.e. *Tscore_cognitive* ) and driver status classification showing the number of drivers in each category using a cut-point value of predicted probability of 0.49.

Using a cut-point of 0.49 on the predicted probability scale, a sensitivity of 50 percent, a specificity of 95.83 percent, a positive predictive value of 66.67 percent, a negative predictive value of 92 percent and a correct classification rate of 89.29 percent was obtained. That is, 89.29 percent of the drivers were correctly classified as either being "poor-drivers" or "not-poor-drivers" by the diagnostic test (*Tscore_cognitive* ). Only two "not-poor-driver" out of 48 were classified as a "poor-driver" by the test and four "poor-drivers" out of 8 as "not-poor-drivers". In total, out of the 56 drivers 6 drivers were misclassified by the test (i.e. *Tscore_cognitive* ). Positive predictive value was relatively not high but negative predictive values of the test was high.

For validating the model, the methodology of Section 8.4 was used. The LOO (leave-one-out) technique was used jointly with the area under ROC (Receiver Operating Characteristic) curve for validating the logistic regression model (Bautista et al., 1999), because typically ROC curves are used to evaluate model discrimination. First, the area under ROC was calculated from the model based on all data and then it was compared with the area under ROC curve obtained from the predicted probabilities using the LOO technique (leave-one-out). The number of covariate patterns in the data were 56, therefore for the LOO technique, the model was fitted 56 times, each time leaving a particular covariate pattern out. The area under ROC curve obtained using all the data was 0.9010 and the area under ROC curve obtained using the LOO technique was 0.8438 (ROC curve and predicted probabilities obtained using LOO are at Appendix-G). The statistical software Stata® was used to test the equality of both areas using a procedure suggested by DeLong et al. (DeLong et al., 1988 cited in Stata® 10 manual), in which the null hypothesis is $H_0$: both areas are equal. Although the area 0.9010 is larger than 0.8438, the chi-squared test yielded a p-value of 0.0548, suggesting that there is no significant difference between the two areas, showing that the discriminatory ability of the logistic regression model did not decrease significantly using the LOO technique and therefore there was reasonably strong evidence that the model would be a satisfactory predictor for determining the status (whether "poor-driver" or "not-poor-driver") of future drivers, based on $Tscore\_cognitive$ .

## 8.7 Results and Discussion

All three scenarios resulted in selection of the best fitting model as the one based on *ufov3*. Also, in all three scenarios, the next best univariate models in decreasing order of over all fit were the models based on *dichotic*, *trail*, *rey-copy* and *paper* cognitive tests. The logistic regression model based on *ufov3* as predictor was:

$$\hat{\pi} = \frac{1}{1+e^{-(-4.591905+0.019135\ ufov3)}} \tag{8.33}$$

Where $\hat{\pi}$ is the predicted probability of a driver being a "poor-driver". Because of the positive sign of the coefficient for *ufov3* in Equation 8.33, the model predicts that the odds of being a "poor-driver" increase with increasing score on *ufov3*. This model reflected that the odds of being a "poor-driver" increase over 2.15 fold for drivers having an increased ufov3 score of 40 compared with the ufov3 scores of other drivers or in other words, the odds of being a "poor-driver" increase 115 percent for drivers having an increased ufov3 score of 40 compared with the ufov3 scores of other drivers. This consistency of the ufov3 test as being a significant predictor is also clear from the fact that it was also one of the predictors defining the best linear regression model selected in Section 7.4.4. The area under ROC (Receiver Operating Characteristic) curve of 0.8659 reflects excellent discrimination (Hosmer and Lemeshow, 2000). According to Kutner et al. (2005), selecting a cut-point value (of predicted probability) of 0.5 is only reasonable when (a) it is equally likely for the outcome of interest (i.e. "poor-drivers") and the complementary outcome (i.e. "not-poor-drivers") to occur in the population of interest; and (b) the cost of incorrectly predicting the outcome of interest (i.e. "poor-drivers") and the complementary outcome (i.e. "not-poor-drivers") are approximately the same. Since these two conditions were not satisfied therefore a cut-point of 0.5 was avoided. As an alternative, Kutner et al. (2005) suggest to use a cut-point value such that the proportion of incorrect predictions is lowest (or the proportion of correct predictions is highest). Therefore, different cut-points were evaluated and finally a value of 0.4 gave the highest correct classification. Using a cut-point of 0.4, a Sensitivity of 62.5 percent, a Specificity of 97.92 percent, a Positive Predictive Value of 83.33 percent, a Negative Predictive Value of 94 percent and an overall correct classification rate of 92.86 percent was obtained. The relatively lower Sensitivity of the test is attributed to the fact that the mean of predicted probabilities of the "poor-driver" group was 0.4652, which was close to the cut-point of 0.4. Only three "poor-drivers" were misclassified as "not-poor-drivers" and one "not-poor-driver" misclassified as a "poor-driver" by the *ufov3* test. Only for comparison purposes with other studies, if the cut point is lowered to 0.2, a Sensitivity and Specificity of 75 and 85.42 percent is obtained respectively, which is quite comparable to the Sensitivity and Specificity of the study by Ball et al. (1993) (Sensitivity= 89%, Specificity= 81%) and Goode et al. (1998) (Sensitivity= 86.3%, Specificity=84.3%) in predicting actual crash history (from state records) of elderly drivers using the UFOV test. Duchek et al. (1998) carried out a study using three groups: (a) healthy control group (58 subjects) (b) very mild DAT (dementia of the Alzheimer's type) group (49 subjects) (c) mild DAT group (29 subjects). All participants were administered the 45 minute Washington University Road Test (WURT). A correlation coefficient of - 0.56 was obtained between the UFOV scores and the driving scores such that poorer driving performance was significantly related to greater reduction in UFOV. In another study of older drivers performed at the University of Alabama at Birmingham, Owsley et al. (1991) found that deficits in information processing ability as measured by the useful field of view test and deficits in cognitive abilities were related to crash involvement as recorded by the state. Owsley et al. (1998b) made an exploratory study (from 1985 to 1990 in the state of Alabama) of 78 drivers involved in injurious crashes, 101 drivers involved in non-injurious crashes and 115 drivers not involved in any crash during the same period. The drivers' age ranged from 55 to 87 years. They reported that elderly drivers having reduction of UFOV greater than 40 percent were at least 20 times more likely to be involved in an injurious crash compared with subjects with no or more minor reductions in UFOV. Studies made by other authors have also found the UFOV test to be predictive of accidents in older drivers (Ball & Owsley, 1993; Rizzo et al., 1997; Sims et al, 1998).

The logistic regression model based on *age* as a predictor had an area under ROC (Receiver Operating Characteristic) curve of 0.9062 which is considered excellent discrimination (Hosmer and Lemeshow, 2000). The model was:

$$\hat{\pi} = \frac{1}{1+e^{-(-12.49311+0.1654765\ age)}} \tag{8.34}$$

Where $\hat{\pi}$ is the predicted probability of a driver being a "poor-driver". Because of the positive sign of the coefficient for *age* in Equation 8.34, the model predicts that the odds of being a "poor-driver" increase with increasing *age*. This model reflected that the odds of being a "poor-driver" increase over 5.23 fold for drivers having an increased *age* of 10 years compared with the *age* of other drivers or in other words, the odds of being a "poor-driver" increase 423 percent for drivers having an increased *age* of 10 compared with the *age* of other drivers. The cut-point value of 0.56 on the predicted probability scale provided an overall correct classification rate of 91.07 percent which was slightly less than that for the *ufov3* model. Using a cut-point of 0.56, a Sensitivity of 50 percent, a Specificity of 97.92 percent, a Positive Predictive Value of 80 percent and a Negative Predictive Value of 92.16 percent was obtained. Four "poor-drivers" were misclassified as "not-poor-drivers" and one "not-poor-driver" misclassified as a "poor-driver" by *age* as a test. The finding that the model with *age* as a predictor had good discrimination was not surprising as Edwards et al. (2005) in their study found out that the UFOV test was significantly correlated with age (and the model based on ufov3 is an excellent discriminator), which was also corroborated by our high correlation coefficient of 0.7231 between age and *ufov3*. The National Highway Traffic Safety Administration report on accident statistics shows a steep rise in per mile automobile accident risk beginning around age 65, with the fatality rate per million miles of travel being 17 times that of the 25-65 age group for those over the age of 65 (NHTSA, 1997 cited in McKnight & McKnight, 1999). Also, past the age of 75, the risk of intersection collisions increases substantially for older drivers in almost all intersection manoeuvres (Staplin & Lyles, 1991; Preusser et al., 1998). It has been found that the time required to search for a target or visual search performance worsens with age (Scialfa et al., 1999 cited in Caird & Hancock, 2002) and spatial orientation declines with the course of normal ageing (Schaie, 1996). This obviously could have implications for intersections and other traffic environments/scenarios, where the traffic is not self-paced and the risk of cognitive overload is high. Older drivers are more sensitive to noise and hence in performing a task, they require stronger signals to react (i.e. they have a lower signal-to-noise ratio) (Lundberg, 2003), thus affecting their performance in cluttered traffic environments etc. Also, the onset of age-associated diseases affecting cognitive functions even further exacerbates the situation and thus increases the risk of motor vehicle crashes in elderly drivers, as these diseases are more prevalent in the older age group. Also, there is some overlap between the extremities between normal ageing and MCI (Mild Cognitive Impairment) and between MCI and between early dementia (Petersen, 2003).

The logistic regression model based on $Tscore\_cognitive$ as a predictor had an area under ROC (Receiver Operating Characteristic) curve of 0.9010 which is considered excellent discrimination (Hosmer and Lemeshow, 2000). The model was:

$$\hat{\pi} = \frac{1}{1 + e^{-(-12.91376 + 0.1067349\ Tscore\_cognitive)}} \tag{8.35}$$

Where $\hat{\pi}$ is the predicted probability of a driver being a "poor-driver". Because of the positive sign of the coefficient for $Tscore\_cognitive$ in Equation 8.35, the model predicts that the odds of being a "poor-driver" increase with increasing score on $Tscore\_cognitive$. This model reflected that the odds of being a "poor-driver" increase over 2.91 fold for drivers having an increased score on $Tscore\_cognitive$ of 10 compared with the score of

$Tscore\_cognitive$ of other drivers or in other words, the odds of being a "poor-driver" increase 191 percent for drivers having an increased score of 10 on $Tscore\_cognitive$ compared with the score of other drivers. The cut-point value of 0.49 on the predicted probability scale provided an overall correct classification rate of 89.29 percent which was slightly less than that for the *age* model. Using a cut-point of 0.49, a Sensitivity of 50 percent, a Specificity of 95.83 percent, a Positive Predictive Value of 66.67 percent and a Negative Predictive Value of 92 percent was obtained. Four "poor-drivers" were misclassified as "not-poor-drivers" and two "not-poor-driver" misclassified as a "poor-driver" by $Tscore\_cognitive$ as a test. Since age related neurodegenerative conditions such as Alzheimer's disease and normal ageing affect multiple but interacting aspects of cognition, a composite index reflecting overall cognitive ability needs to be evaluated for the prediction of driving safety (Anderson et al., 2005). Therefore, the model consisting of a predictor based on the composite of all nine cognitive tests ($Tscore\_cognitive$) was explored as per recommendation of Anderson et al. (2005). However, it is apparent that the discriminatory power of this model is not better (but not too worse) than the logistic model based on the single *ufov3* test, showing that identification of drivers exhibiting risky driving behavior due to cognitive impairment can be brought about better by employing *ufov3* test than by using a composite of nine cognitive tests (*trail, clock, rey-copy, rey-recall, dichotic, paper, ufov1, ufov2, ufov3*). The Anderson et al. (2005) study included a simulation drive with a sample of 202 older drivers (ages 55 and older) out of which 70 drivers had mild dementia. Their findings supported the concept of a composite cognitive index, however it may be noted that their battery of tests did not include the UFOV test (although *rey-copy*, *rey-recall* and *trail* were common to both batteries) and a significant proportion of their sample was in a state of mild dementia (which could have made effects appear stronger).

## 8.8 Summary

This chapter has described the development of univariate logistic regression models using "poor-drivers" and "not-poor-drivers" as the two categories of the dependent variable and each of the nine cognitive tests as a candidate for being a predictor. The categorization of the three driver groups into two driver groups ("poor-drivers" and "not-poor-drivers"), based on driving performance measures has been explained. The use of univariate logistic regression rather than multiple linear regression has been justified based on the sample size, lopsidedness of the dependent variable and the number of potential candidate predictors. The process of development of univariate logistic regression models for each of the three scenarios (i.e. *pessimistic*, *optimistic* and *deletion of driver "O56"* scenarios) has been explained, and the selection of the best/parsimonious model has been demonstrated. All three scenarios resulted in selection of the best model as the one based on *ufov3* as a predictor. Also, in all three scenarios, the next best univariate models in decreasing order of over all fit were the models based on *dichotic*, *trail*, *rey-copy* and *paper* cognitive tests. Univariate regression model based on *ufov3* was developed, its diagnostics performed and validated using the LOO technique. In order to determine the extent to which age alone could discriminate between "poor-drivers" and "not-poor-drivers", a logistic regression model based on *age* was developed, its diagnostics performed and validated using the LOO (leave-one-out) technique. To further explore if a cognitive measure that is based on a composite score of all nine (*trail, clock, rey-copy, rey-recall, dichotic, paper, ufov1, ufov2, ufov3*) neuropsychological tests was a better discriminator than the *ufov3* test alone, a univariate logistic regression model based on such a cognitive measure ( i.e. $Tscore\_cognitive$ ) was developed, its diagnostics performed and validated using the LOO technique. Based on the area under ROC curve, all three models were considered to have excellent discrimination. However, the logistic regression model based on *ufov3* test had the best discrimination.

# 9 Conclusions

## 9.1 Introduction

The proportion of licensed drivers is increasing in the general driving population and a substantial number within this population group are experiencing a cognitive decline in functions that are critical to the driving task. According to one estimate, about 40 percent of the driving population will be over the age of 60 by the year 2020 in the UK and currently, several hundred thousand drivers with dementia hold driving licenses (Groeger, 2000). It is well known that the number of motor vehicle crashes per unit distance of automobile travel is "U"-shaped, with risk increasing slightly between the ages of 55 and 60, but greater increases in risk with each successive five-year interval. The fatality rate of drivers over the age of 80 is even higher than that of drivers less than 24 years of age. With regard to issue of neuropsychological tests, there is little consensus on which tests can be used to predict driving safety; therefore, there is no standard testing protocol (that is reliable) for assessing a person's fitness to drive after the onset of neurological disease/trauma and/or natural ageing. In the absence of a reliable standard protocol, some clinicians make their judgments based on self-report (of drivers), which has risks associated with it as lack of insight and judgment are potential common traits of the population experiencing cognitive decrements. Therefore the decisions regarding fitness to drive exude a low level of confidence on part of the clinicians/professionals. Seldom is recourse made by health professionals to on-road driving assessment as a first alternative as it requires a fee and such testing centres are not readily available everywhere. Thus there is a need for more information on assessment on fitness to drive (with regard to cognitive tests), since medical information alone is not sufficient to assist in decision making. This will also alleviate the need for the requirement of an on-road evaluation /assessment or can be a supplementary tool in addition to on-road assessment and will instill more confidence in decision making.

The difference between the direct effects of normal ageing and that of abnormal ageing (dementing disease especially in the early stage) relevant to driving skills, is less than clear-cut. It is possible that subjects exhibiting subtle cognitive changes may in fact have transgressed into the early stage of Dementia. Some individuals who have mild dementia possess sufficient driving skills to be designated as fit drivers; however, a stage /time will come when their cognitive impairment will exacerbate and will ultimately render them unfit drivers. The most challenging assessment and decision for the physician / licensing authority as regards fitness to drive lies in drivers who are questionably demented or are in a state of very mild dementia. This research addresses this issue of the identification of cognitive (neuropsychological) tests that can be used to assess an individual's ability to drive and especially of those individuals that are questionably demented and are the most difficult to identify. Nine cognitive tests (*trail, clock, rey-copy, rey-recall, dichotic, paper, ufov1, ufov2, ufov3*), where the *rey-copy* and *rey-recall* were essentially parts of one test (Rey-Osterrieth Complex Figure Test) and *ufov1, ufov2* and *ufov3* were the three subsets of the UFOV test, were administered to a sample of 56 individuals comprising of two age groups (young and old). Driving performance was gauged through two simulated drives on the STISIM driving simulator as it provided the most appropriate means of assessing/identifying risky driving behavior due to cognitive impairment, because in a driving simulator, drives can be designed based on psychometric principles and also, each driver is subjected to the same opportunities for committing errors.

In this research, a series of cognitive tests have been identified suitable to be used to assess an individual's ability to drive and especially of those individual's that are questionably demented or are in a subtle state of cognitive decline (and therefore are the most difficult to identify). The detail of these cognitive tests is provided in Chapter 3 and the mechanisms describing their selection is laid out in Chapters 7 and 8. Out of the 56 drivers, 8 drivers (all from the older age group) possessing "poor-driving" skills and/or exhibiting risky driving behavior were identified by the simultaneous use of more than one variable through the novel approach of normal-mixture-model cluster analysis. In addition to studying the relative ability of cognitive tests to discriminate between "poor-driving" and "not-poor-driving", the effect of other factors such as age and a composite cognitive measure (based on all nine cognitive tests) was also evaluated in discriminating "poor-drivers" and "not-poor-drivers".

## 9.2 Conclusions

- A major contribution by this research has been the development and deployment of a methodology by which drivers can be categorized including a deficient/poor driver group. The categorization is based on driving performance skills by considering more than one (i.e. three) driving performance parameters simultaneously using: (1) a diverse sample of subjects, (2) simulation drive-designs based on specific psychometric principles, (3) calculation of driving performance indices by removing parameters contributing to "noise" and keeping the ones contributing to "signal" through the concept of Cronbach's reliability coefficient and weighting, and (4) the technique of normal-mixture-model cluster analysis. This has led to the identification of a deficient/poor driver group with some individuals even exhibiting relatively subtle changes in their driving performance. Various other studies suffer from methodological deficiencies due to one or more of the following: (1) employment of inappropriate design of simulated drives, (2) selection of clinical samples, (3) use of a single driving performance parameter/index for driver classification and using the normal group as a reference for decision making, (4) using relatively fewer driving performance parameters for gauging driving performance, and (5) inappropriate choice of cognitive tests etc.
- A straightforward cognitive test has been identified which can discriminate between "poor" and "not-poor" drivers. The main finding and contribution from this study with regard to the relative discriminating ability of different cognitive tests is that the *ufov3* test fared the best and showed the highest discriminating ability in separating "poor drivers" from "not-poor-drivers". The *ufov3* test is the third subset of the UFOV (Useful Field of View) test and takes up the most cognitive effort compared to *ufov1* and *ufov2*. This test cannot be administered in isolation, because the examinee has to get gradually accustomed to the difficulty of the task by first familiarizing himself with *ufov1* and *ufov2*. The next best discriminating ability in decreasing order of strength is that of *dichotic*, *trail*, *rey-copy* and *paper*. This highlights the relevance of visuospatial skills and attention

in gauging risky driving behavior, as the UFOV test primarily evaluates visual processing speed and divided and spatial attention. The cut-point value of predicted probability of 0.4 adopted in Section 8.4 when used in logistic regression Equation 8.33 corresponds to a cut-point value of 219 (after rounding) of *ufov3*. At this threshold, three "poor-drivers" and one "not-poor-driver"' were misclassified (i.e. false negatives=3, false positives=1). Lowering this threshold although would have increased Sensitivity but at the expense of lowering Specificity. The test had moderate Sensitivity when compared to its Specificity. Although high Sensitivity for a test is important, but an effective screening tool should not also incorrectly identify someone who is able to drive competently as a highly *specific* test is rarely positive in the absence of the outcome of interest (i.e. being a "poor-driver") i.e. it gives few false positive results and thus a large number of "not-poor-drivers" will not be referred to on-road tests, which is also quite favourable as these "not-poor-drivers" will be saved undue emotional distress and expense. For clinicians who practice in areas where driver evaluation resources (i.e. on-road test facility) are not available and who must decide to remove driving privileges based on clinical findings, this is an important consideration. It may be noted that our sample consisted of active drivers from the general driving population (not a clinical population), who apparently had good mental and physical constitution (and therefore the differences between the younger group and older group were much subtler). Had the older group come from a clinical population, the effects would have appeared stronger with consequent higher Sensitivity for the test. Some misclassification by using the test is bound to happen no matter how we ascertain the cut-point because of the overlap of the distributions of the "poor-drivers" and the "not-poor-drivers" i.e. some "poor-drivers" will tend to have *ufov3* scores lower than "not-poor-drivers" while some "not-poor-drivers" will have *ufov3* scores higher than "poor-drivers". This compromise will happen because we are using a *simpler test* (i.e. the *ufov3* test) as a proxy for a more elaborate, time consuming, expensive and accurate test (i.e. the simulator test) for ascertaining "poor-drivers" with the understanding that some misclassification will result. This risk is justified due to the convenience and the low error rate of classification of the *ufov3* test (or the high rate of correct classification of the *ufov3* test, keeping in view that no cognitive measure was used in driver-categorization). Also, in practice, the test is most likely to be applied to a population / individuals exhibiting not-so-subtle cognitive decrements, which is bound to give much higher sensitivities.

- Age has excellent discrimination ability in separating "poor drivers" from "not-poor-drivers". This discriminating ability is slightly less than that of the *ufov3* test, which is evident from the relatively smaller Sensitivity of the age model, although in itself the discrimination achieved with *age* as a predictor is considered excellent, based on area under ROC curve. Using age as a discriminator, four "poor-drivers" and one "not-poor-driver"' were misclassified. The cut-point value of predicted probability of 0.56 adopted in Section 8.5 when used in Equation 8.34 corresponds to a cut-point value of 77 years (after rounding) of *age*. However, age cannot solely be made as a criterion for the discrimination of drivers as its effects are confounded by neurological diseases. In this research, the factor *age* may have emerged as a significant discriminator because its effects were not confounded by other diseases as the sample was not from a clinical population or had individuals of diverse age afflicted with neurological ailments. The elderly are considered a very heterogeneous group. Even in older people of the same age, there exists considerable variability in different attributes. After middle age, health deteriorates exponentially over a period of 1 to 3 decades. Some older people experience a quite rapid decline in health while others will have a slow decline and they will be afflicted with disabilities quite late in life. Hence, a global ruling cannot be made in this regard. Also, certain medical conditions / neurological diseases that have a tendency to bring about cognitive impairment to the extent that safe operation of motor vehicles is not possible and increasingly, such medical conditions have started to afflict people at relatively early ages. Even a small but significant number of younger people suffer from dementia who are likely to drive a motor car. This implies that the government should take age into account in formulating policies that promote on-road safety with regard to the mandatory evaluation of older driver performance and/or cognitive evaluation.

- A composite cognitive measure that is based on all nine neuropsychological tests (*trail, clock, rey-copy, rey-recall, dichotic, paper, ufov1, ufov2, ufov3*), is not a better discriminator than the single *ufov3* test in separating "poor drivers" from "not-poor-drivers". This composite cognitive measure was the sum of nine cognitive tests after they had been standardized and four of them oriented (standardized scores of *clock*, *rey-copy*, *rey-recall* and *paper* multiplied by -1) in the proper direction. Although in itself the discrimination achieved with this composite cognitive measure based on area under ROC curve is considered excellent. This finding is contrary to some studies. It is possible for this research to have come with this finding because of the particular methodology/approach and battery of cognitive tests adopted. Using this composite cognitive measure as a discriminator, four "poor-drivers" and two "not-poor-driver"' were misclassified. Even *age* as a discriminator is slightly better in performance than this composite cognitive measure. It shows that the *ufov3* test is tapping relevant cognitive constructs (with regard to driver discrimination) than the "test all" cognitive skills approach that is being exercised through the composite cognitive measure. Also, this is testament to the much needed parsimony/economy in cognitive testing for driving and implies that preliminary driving-status can be determined without extensive investment in time.
- An equation to predict General driving performance through an index has been derived. The cognitive tests *ufov3*, *dichotic* and *rey-recall* as a group emerged as the best predictors of a *general driving skills* index in this research. This index is a measure of the general driving skill of a driver with essentially the same emphasis being placed on each driving performance parameter and therefore cannot be used to assess risky driving behavior due to cognitive impairment. However, it is a useful general index that can be used to gauge driving proficiency. These three tests as a whole measure visual-spatial construction ability, visual memory, organizational, planning and problem solving skills (executive functioning), motor functioning, ability to switch attention, vulnerability to distraction,

selective attention, visual processing speed, divided and spatial attention —— traits that are crucial for driving proficiency. This index will enable to predict general driving skill without putting an individual to a road-test and will show the degree of aptitude in handling a vehicle in diverse maneuvers.

- The clock drawing test (*clock*) and the trail making test (*trail*) often used to clinically assesss dementure did not emerge as significant predictors of driving ability. Both tests are quick and easy to administer (are paper-and-pencil tests) and are recommended by the American Medical Association (AMA) for screening unsafe drivers. However, studies (Powlishta et al., 2002; Lee et al., 1996; Storey et al., 2001) have shown the clock drawing test not to be a good screening instrument for detecting the very earliest signs of dementia. Similarly, a critique of the trail test in this regard has been highlighted in Section 7.4.8.

## 9.3 Recommendations

This research shows that relatively simple and inexpensive neuropsychological tests of specific cognitive abilities can be used to aid the evaluation of drivers' risk of unsafe driving and measurement of general driving skill. Scores greater than the threshold scores on the *ufov3* test (that follow) should not result in cessation of driving privileges, but rather indicate that the driver needs further assessment / evaluation. This evaluation could mean the assessment of the driver by a qualified and experienced driving evaluation specialist / instructor, before a final decision is taken. Also, this cognitive test (*ufov3*) is not a substitute for other tests/examinations that ascertain physical and sensory (e.g. vision, hearing etc) soundness of drivers. Also, clinicians should expect driving skills to deteriorate over time. Therefore, planning for repeat testing after an appropriate time interval is necessary in the very old drivers / critical cases. As a whole, it is expected that if the following recommendations are followed, they can result in substantial safety for older drivers and drivers in general.

- A score on *ufov3* test greater than 220 indicates that the driver should be further evaluated by a driving specialist to ascertain risky driving behaviour.

- Drivers above the age of 77 pose a potentially greater risk. If such drivers have ufov3 scores greater than 220 it indicates much stronger evidence for the necessity to ascertain risky driving behavior through a driving specialist.
- The UFOV test should be made mandatory for drivers above the age of 77.
- If the intent is to measure *general driving skill,* then the cognitive tests *ufov3*, *dichotic* and *rey-recall* can be used with the following equation (Equation 7.13) for prediction purposes:

  $$DPI = 104.31 + 23.40\left(dichotic + 1\right)^{-0.5} - 7.83 \log_e\left(ufov3\right) + 0.72\left(rey - recall\right)$$

  As a guide, predicted scores less 96 (i.e. 2 standard deviations below the mean of the younger group) would signify poor driving proficiency and greater than 110 (median of the younger group) would signify good driving proficiency.

It may be reiterated that the *ufov3* test (involves the identification of a central target and simultaneously the radial localization of a peripheral target embedded in distracter triangles) is the third subtest of the UFOV (Useful Field of View) test. The UFOV test takes about 15 minutes to administer using an ordinary on-the-market personal computer with a 17 inch monitor (computer literacy is not a prerequisite for taking the test).

## 9.4 Recommended Future Research

There is considerable scope for further research in the realm of neuropsychological testing (cognitive testing). One aspect that clearly emerges from this research is that in general, the paper-and-pencil tests did not emerge as significant predictors of driving skill and in discriminating “poor drivers” from “not-poor drivers”. The computerized/listening tests were more promising. In this context, it is perceived that the ufov3 test emerged as a significant predictor of driving skill and in discriminating “poor drivers” from “not-poor drivers” because in the ufov test, the stimulus presentation time is shortened if the previous two responses are correct and is lengthened if the previous response is incorrect. This process of tracking the perceptual threshold is continued until a stable estimate of 75 percent correct is calculated. Therefore, new/future neuropsychological tests should dynamically adapt to the performance of an individual in setting threshold values so that ceiling effects (too many individual achievieng the highest score) and floor effects (too many individuals achieving the lowest score) are avoided so that discrimination between the individuals is obvious and variability in scores is achieved with scores proportional to an individual's performance/ability on the test; this is only possible though computerized tests. These test should tap different aspects of attention and diverse visuospatial abilities (as these are the most important cognitive attributes with regard to driving) and decision making abilities under time pressure. Also, much larger samples of drivers could be employed from the general driving population in future research.

# Glossary

<u>Automatic Processes :</u> Driving performance also depends on whether individual skills of driving draw on "automatic" or "effortful" processes (controlled processes). "Automatic processes" are fast, involuntary, and place limited demands on attentional capacity. Automaticity generally develops under highly predictable stimulus conditions. In contrast, "effortful" or "controlled processes" are slow and capacity-demanding and are used to deal with unpredictable or unfamiliar stimulus demands. Some driving tasks become automatic following extensive practice (e.g., shifting gears). Other may continue to require effortful processing even in highly skilled drivers (e.g., backing into a street from a driveway). (Parasuraman & Nestor, 1993).

<u>CDR (Clinical Dementia Rating):</u> is a more global assessment tool that incorporates both cognitive and functional assessment. It is used by researchers as a mental status screen and is recommended in the consensus guidelines of many professional organisations. It is a five point scale and is use to rate function in six cognitive functional categories. A rating of 0.5 indicates questionable impairment, 3 indicates sever impairment. This paradigm is unlikely to be used outside the field of research or in specialized dementia centers because of the training and time required to administer the test.

<u>Effortful Processes (Controlled Processes):</u> Driving performance also depends on whether individual skills of driving draw on "automatic" or "effortful" processes (controlled processes). "Automatic processes" are fast, involuntary, and place limited demands on attentional capacity. Automaticity generally develops under highly predictable stimulus conditions. In contrast, "effortful" or "controlled processes" are slow and capacity-demanding and are used to deal with unpredictable or unfamiliar stimulus demands. Some driving tasks become automatic following extensive practice (e.g., shifting gears). Other may continue to require effortful processing even in highly skilled drivers (e.g., backing into a street from a driveway). (Parasuraman & Nestor, 1993).

Executive function: Executive function is considered to be a product of the coordinated operation of various processes to accomplish a particular goal in a flexible manner. The mechanism or system responsible for the coordinated operation of various processes is called executive control. (Funahashi, 2001).

Selective Attention: Selective attention refers to the preferential processing by the brain of particular stimulus events. It involves the focusing and shifting of attention between stimulus locations, features, or categories and is generally evaluated by tests such as dichotic listening, visual search, cue-directed detection, and the Stroop test. (Parasuraman & Nestor, 1993).

Sensitivity: Is the ability of a screening test to identify those individuals who have a condition under consideration. It is calculated as the percentage of all cases having the condition (e.g. the condition could be passing the driving test or having a particular disease) who are judged by the test to have the condition (e.g. the condition could be passing the driving test or having a particular disease). Few real cases are missed by tests that have a high sensitivity and so will generate few "false-negative" judgements. Sensitivity = (True Positive)/ (True Positive + False Negative). Sensitivity alone does not tell us all about the test, because a 100 percent sensitivity can be trivially achieved by labelling all test cases positive; hence, we also need to know the Specificity of the test (McDowell, 2006).

Specificity: Is the ability of a screening test to correctly identify those individuals who do not have the condition under consideration. It is calculated as the percentage of all cases not having the condition (e.g. the condition could be passing the driving test or having a particular disease) who are judged by the test not to have the condition (e.g. the condition could be passing the driving test or having a particular disease). Specificity = (True Negative)/ (True Negative + False Positive). (McDowell, 2006).

Divided Attention: Divided Attention is involved when two or more stimulus sources must be monitored or when two or more tasks have to be performed simultaneously. (Parasuraman & Nestor, 1993).

Sustained Attention: Sustained attention refers to the maintenance of an alert state for prolonged periods of time and is typically measured in vigilance tasks. (Parasuraman & Nestor, 1993).

Working Memory: Working memory is the collection of structures and processes within the brain used for temporarily storing and manipulating information. In other words, the major function of working memory is to hold several interrelated bits of information in our mind, all at the same time , so that this information can be worked with and then used. Working memory acts like a work bench, where material undergoes handling, combination and transformation. New material and old material, both are held by this workbench which has been retrieved from storage (long-term memory).The four main structures of the working memory model are the Central Executive, the Visuospatial Sketchpad, the Phonological loop and the Episodic buffer, as shown below in Figure G.1. The Central Executive integrates information from the Phonological loop, the Visuospatial Sketchpad and the episodic buffer. In attention, planning strategies and coordinating behaviour, the Central Executive plays a major role. It suppresses irrelevant information and distributes tasks into appropriate areas and assigns priority and mental capacity to whichever task is seen more important than the other. The central executive can be metaphorically thought of as an executive supervisor of a firm/organisation. Issues that need to be attended to and issues that need to be ignored are decided by the executive. Also, figuring out how to tackle a problem by the selection of strategies is the work of the executive. The Central Executive has a limited ability to perform simultaneous tasks. The Visuospatial Sketchpad holds images and visual data and from that manipulates and processes the stimuli in order to produce the desired outcome, e.g. in driving, judging distance uses this function. The Phonological loop deals with the manipulation and retention of auditory data, such as a particular sounds or words, a practical use for this is rehearsing a phone number while trying to find a piece of paper to write it on so you don't forget. The episodic buffer acts as a temporary storehouse, where information from the phonological loop, the visuospatial sketchpad and long-term memory is gathered and combined. In order to interpret an earlier experience, solve new problems and plan future activities, the episodic buffer actively manipulates information (Matlin, 2005; Baddeley, 2000).

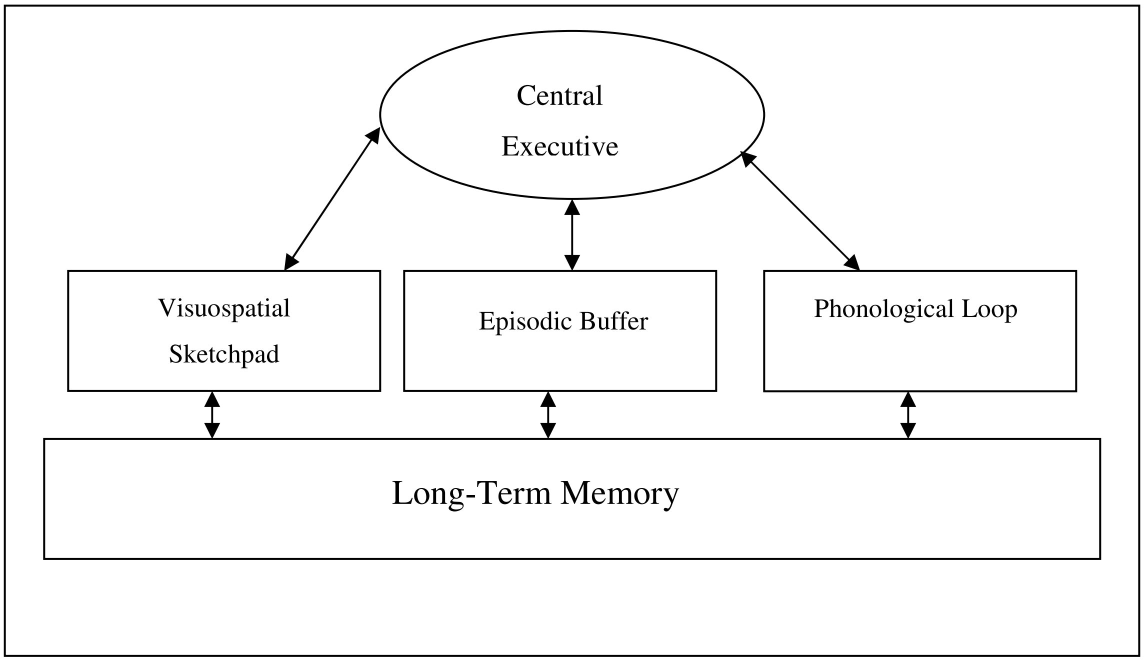



**Figure G.1** Model of Working Memory showing the Phonological loop, the Visuospatial Sketchpad, the Central Executive and Episodic Buffer—as well as their interaction with Long-term Memory (Baddeley, 2000 cited in Matlin, 2005)

# Appendix-A

## <u>Instruction Sheet for Trail Making-B Test</u>

### Sample Instructions

Start with the sample side up of the Trail Making Test Part B. Provide the examinee with a pencil and the following instructions:

- On this page are some numbers and letters (point to them). Start at number 1 (point to "1") and draw a line to A (point to "A"), A to 2 (point to "2"), 2 to B (point to "B"), B to 3 (point to "3"), 3 to C (point to "C"), and so on until you reach the end (point to circle marked "END"). Remember first you have a number (point to "1") and then a letter (point to "A"), then a number (point to "2"), then a letter (point to "B"), and so on. Draw the lines as fast as you can without lifting the pencil from the paper. Ready! Go!
- If the participant completes the sample completely, say: "That was good, now we will go to the next one." And then proceed to the formal Trail Making-B Test.
- If a mistake is made by the participant (in the sample test), it should be pointed out and explained. The following explanations of mistakes are to be accepted.
- You started with the wrong circle, this is where you start (point to "1").
- You omitted this circle (i.e. missed it). You should go from point 1 to A (point), A to 2 (point), 2 to B (point), B to 3 (point), and so on until you reach the end point. After the mistake has been explained, the examiner should mark out the wrong part and say: "Go on from here (pointing to the last circle completed correctly in the sequence)." If it is obvious that the examinee intended to touch the circle but missed it, it should not be counted as an omission, but he or she should be cautioned to touch the circle.
- If the examinee still cannot complete the sample test, then his hand should be held and the pencil guided (eraser side down) through the circles. Then say: "Now try

it, remember that start at number 1 (point) and a line should be drawn from 1 to A (Point), from A to 2 (point), 2 to B (point), from B to 3 (point), and so on until you reach the END. I will time how fast you can do this. Ready? Start!"

- If the examinee succeeds this time, then go on to the formal Trail Making-B Test. If he doesn't succeed then the procedure should be repeated until he succeeds or it becomes evident that he cannot.

## Formal Trail Making-B Test Instructions

On this page are both numbers and letters. It should be done the same way as the sample was done. The examinee should be instructed to begin at number1 (point) and a line drawn from 1 to A (point), from A to 2 (point), 2 to B (point), B to 3 (point), 3 to C (point), and so on ,in order until the END is reached. It should be reiterated that first there is a number (point), and then a letter (point), then a number (point), then a letter (point), and so on. The examinee should be told not to skip around and should go from one circle to the next in appropriate order. And the lines should be drawn as fast as possible without lifting the pencil from the paper. Finally say: "Ready! Begin!"

Start timing as soon as the examinee is told to begin. If the participant makes an error, it should be called to his attention and have him proceed from the point the mistake occurred. Do not stop timing. If part B is completed without errors, record the time to the nearest second and remove the test sheet.

Again if it is obvious that the examinee will be unable to complete the task successfully in any amount of time, the activity should be discontinued.

Figure A.1 shows the full version of the Trail Making-B Test consisting of 25 circles with numbers 1 to 13 and alphabets A to L and Figure A.2 the sample test version of the Trail Making-B Test consisting of 8 circles with numbers 1 to 4 and alphabets A to D.

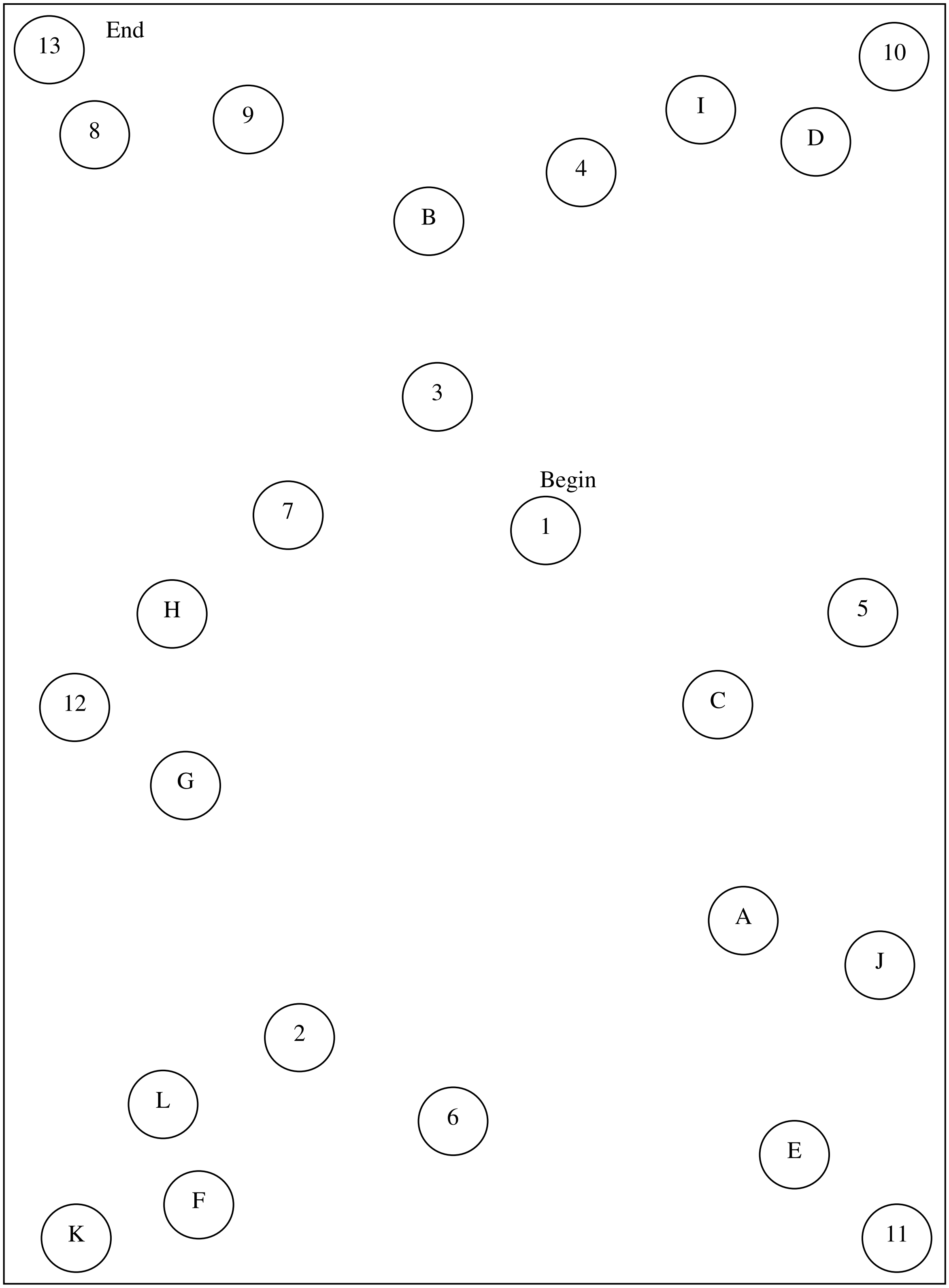



**Figure A.1** Full version of the Trail Making-B Test consisting of 25 circles with numbers 1 to 13 and alphabets A to L.

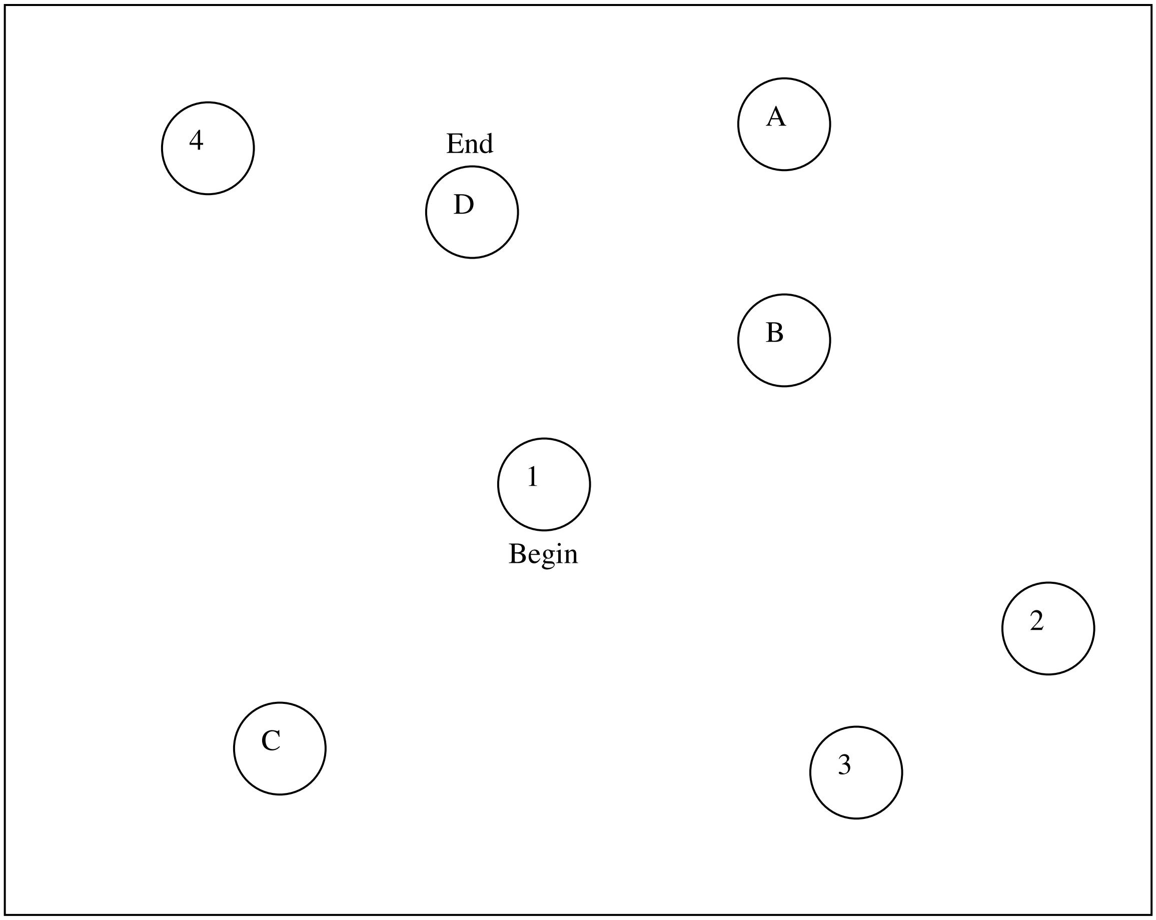



**Figure A.2** Sample test version of the Trail Making-B Test consisting of 8 circles with numbers 1 to 4 and alphabets A to D.

## Instruction Sheet for Clock-Drawing Test

Present the subjects with a blank sheet of paper measuring 8 ½-by-11 inches and the following instructions should be given: "You are to draw a clock and put in the numbers." After the participant draws the clock, the following instructions should be passed: "Now set the time at 10 minutes past 11". Figure A.3 shows the scoring sheet for the Clock Drawing Test (Freedman Method).

| Clock Drawing Test Scoring Sheet (Freedman Method) | |
|---|---|
| Contour | |
| 1. Acceptable contour drawn | |
| 2. Contour is not too small nor overdrawn nor reproduced repeatedly | |
| Numbers | |
| 3. Only numbers 1-12 (without adding extra numbers or omitting numbers) | |
| 4. Arabic number representation | |
| 5. Numbers written in the correct order | |
| 6. Paper not rotated while drawing numbers | |
| 7. Numbers in the correct position | |
| 8. All numbers located inside contour | |
| Hands | |
| 9. Clock has two hands / or marks | |
| 10. Hour target number indicated in some manner | |
| 11. Minute target number indicated in some manner | |
| 12. Hands in correct proportion (minute hand longer) | |
| 13. No superfluous markings | |
| 14. Hands are joined or within 12 mm (1/2 in) of joining | |
| Center | |
| 15. Clock has a center (drawn or inferred/ extrapolated at the point where 2 hands meet) | |
| Total | |

**Figure A.3** Scoring Sheet for Clock Drawing Test (Freedman Method)

Notes: In this scoring system of the free-drawn clock, the clock drawings are scored according to the broad categories of contour, numbers, hands, and center. The scoring system consists of 15 critical items that constitute a maximum score of 15 i.e  a score of 1 for each item if it is in the affirmative. With regard to item 1, the definition of "acceptable" is any closed contour and closure is considered adequate if the lines used to draw the contour are touching or overlapping. If all numbers and hands can not be included in the contour, the size of the contour is considered unacceptable. Roman numerals are also acceptable. If a number is placed in a position that is not normally occupied by another number, then it is considered to be placed correctly. With regard to item 12, if the hour hand is perceptibly or measurably shorter than the minute hand, the hands are considered to be proportioned correctly. If the hands do not physically touch or intersect, the subject's hands (i.e clock hands) can be extended to a point of intersection so as to give an approximate extrapolated center. Superfluous markings are those markings that are not necessary for the indication of contour, numbers, or time. Marks that distinctly represent either 1- or 5-minute intervals are not considered superfluous.

## Instruction Sheet for Rey-Osterrieth Complex Figure Test

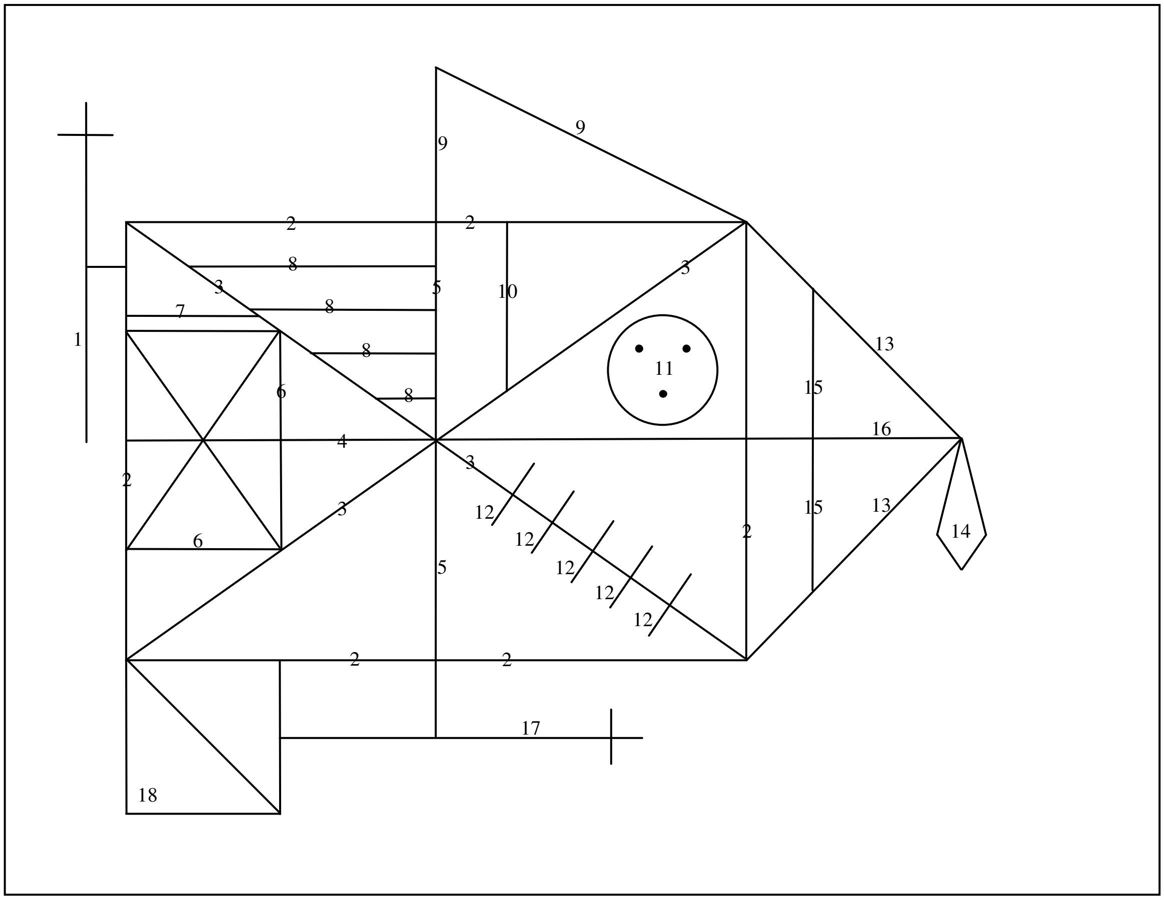
These instructions for the Rey-Osterrieth Complex Figure Test have been adapted from Strauss et al. (2006) (Strauss et al., 2006). Give the candidate a sheet of plain 8 ½ ×11 inches paper which should be placed vertically on the table. Then give the following instructions: I would like to give you a card on which there is a design and I would like you to copy the design on the blank sheet of paper. Please copy the figure with care. The timing should be begun as soon as the drawing is exposed. Erasing is allowed by Meyers and Meyers (Meyers & Meyers, 1995b cited in Strauss et al., 2006). The drawing should be carefully supervised, especially in the early stages. If the candidate is drawing carelessly, he (she) should be reminded that the copy has to be made as accurate as possible.

The maximum and minimum exposure time for the card and the candidate's copy is 5 minutes and 2½ minutes respectively. If after a lapse of 2 ½ minutes it is felt that the candidate is drawing too slowly, he (she) should be made aware of this and told to speed up the drawing. If the drawing is finished before 2 ½ minutes, the candidate should be told to check the drawing carefully and to make sure that it is complete. At the completion of the drawing, the subject's copy as well as the stimulus card (card with the drawing of the Rey-Complex figure) is removed from the candidate's sight.

The drawing completion time should be recorded. It should not take more than 5 minutes to complete the drawing, unless the candidate has a considerable motor difficulty. More important is the fact that the candidate completes the drawing as well as he can, than to have it finished within 5 minutes; therefore, the candidate should be allowed as much time as needed to enable him to draw to the best of his ability. But the card has to be exposed for a maximum of 5 minutes and a minimum of 2 ½ minutes.

After a delay of 3 minutes (after copying the figure), which should be filled with a verbal task such as talking (no visual memory task should be given in this gap), a clean sheet of paper should be provided to the candidate and the following instructions given: If you remember, I gave you a figure to copy a short time ago. I would like that you draw that figure again. There is no time constraint.

The quantitative scoring criteria of the Rey-Osterrieth Test is shown in Figure A.4 below.




| REY-OSTERRIETH COMPLEX FIGURE TEST<br>FORM A (Rey Figure) | | |
|---|---|---|
| Details | Copy | 3-Minutes Recall |
| 1. Cross upper left corner, outside of rectangle | | |
| 2. Large rectangle | | |
| 3. Diagonal cross | | |
| 4. Horizontal midline of 2 | | |
| 5. Vertical midline | | |
| 6. Small rectangle, within 2 to the left | | |
| 7. Small segment above 6 | | |
| 8. Four parallel lines within 2, upper left | | |
| 9. Triangle above 2 upper right | | |
| 10. Small vertical line within 2, below 9 | | |
| 11. Circle with three dots within 2 | | |

| REY-OSTERRIETH COMPLEX FIGURE TEST<br>FORM A (Rey Figure) | | |
|---|---|---|
| 12. Five parallel lines with 2 crossing 3, lower right | | |
| 13. Sides of triangle attached to 2 on right | | |
| 14. Diamond attached to 13 | | |
| 15. Vertical line within triangle 13 parallel to right vertical of 2 | | |
| 16. Horizontal line within 13, continuing 4 to right | | |
| 17. Cross attached to low center | | |
| 18. Square attached to 2, lower left | | |
| Total score | | |

Scoring:

Consider each of the eighteen units separately, and appraise accuracy of each unit and relative position within the whole of the design. For each unit count as follows:

| | | |
|---|---|---|
| Correct | placed properly | 2 points |
| | placed poorly | 1 point |
| Distorted or incomplete | placed properly | 1 point |
| but recognizable | placed poorly | ½ point |
| Absent or not recognizable | | 0 points |
| Maximum | | 36 points |

**Figure A.4** Rey-Osterrieth Complex Figure Test: Form A and legend. Source: Osterrieth, P.A. (1944). Le test de copie d'une figure complex: Contribution a l'etude de la perception et de la memoire. Archives de Psychologie;30: 286-356.

Taylor Scoring Criteria / guidelines for the Rey-Osterieth Complex figure (reproduced from Spreen & Strauss, 1991).
(Note: Detail 1 means Item 1, Detail 2 means Item 2, and so on)

Detail 1: The cross at the upper left corner, outside of the rectangle. The cross must come down to the horizontal midline of the rectangle and must extend above the rectangle. The line that joins the cross to the rectangle must be approximately in the middle of the cross and must come between Detail 7 and the top of the rectangle.

Detail 2: The large rectangle. The horizontal dimensions of the rectangle must not be greater than twice the vertical dimensions of the rectangle, nor must the rectangle resemble a square. Because there are so many possibilities of distorting the rectangle and it is not possible to score for position, a score of ½ point is given if the rectangle is incomplete or distorted in any way.

Detail 3: The diagonal cross must touch each of the four corners of the rectangle and must intersect in the middle of the rectangle.

Detail 4: The horizontal midline of the rectangle must go clearly across from the midpoint of the left side of the rectangle to the midpoint of the right side of the rectangle in one unbroken line.

Detail 5: The vertical midline must start at the midpoint of the bottom of the rectangle and go through in one unbroken line to the midpoint at the top of the rectangle. In scoring for position of details 4, 5, and 6, they should intersect at the midpoint of the rectangle. Usually, if they do not, only one of them is scored as incorrect for position. Very seldom are all three scored as incorrect for not being in position.

Detail 6: The small rectangle within the large rectangle and to the left side of it. The boundaries of Detail 6 are defined by the top of the rectangle falling between lines 2 and 3 of the parallel lines that make up Detail 8, and the width of the small rectangle must be approximately one-quarter of the width of the large rectangle; that is, it should come to the midpoint between the left side of the large rectangle and the vertical midpoint of the rectangle. The cross within detail 6 must come from the four corners of the rectangle and should intersect at the mid-point of the rectangle (i.e., intersecting on Detail 4).

Detail 7: The Straight line above Detail 6 must be shorter than the horizontal aspect of Detail 6 and must fall between the top of Detail 6 and the second line of Detail 8.

Detail 8: The four parallel lines within the rectangle in the upper left corner should be parallel, with the spaces between them approximately equal. If the lines are unduly slanted or, of course, if there are more or less than four of them, then the scoring is penalized.

Detail 9: The triangle above the rectangle on the upper right, with the height less than the base.

Detail 10: The small vertical line within the rectangle just below detail 9. The line should be clearly shifted to the left within the upper right quadrangle in the rectangle.

Detail 11: The circle with three dots must be in the lower right half of the upper right quadrangle. It must not touch any of the three sides of the triangular area in which it is placed, and the positioning of the dots should be such that there are two above and one below, so that it resembles a face.

Detail 12: The five parallel lines that are crossing the lower right aspect of Detail 3 must all be within the lower right quadrangle. They must not touch any sides of the quadrangle, and they should be approximately equidistant from one another.

Detail 13: The triangle on the right end of the large rectangle. The height of the triangle must not be greater than half of the horizontal midline of the rectangle and, as already mentioned, the slope of the sides of the triangle must not be a continuation of the slope of Detail 9.

Detail 14: The diamond attached to the end of Detail 13 should be diamond-shaped and must be attached to the end of detail 13; it must not extend down below the bottom of the large rectangle, Detail 2.

Detail 15: The vertical line within triangle 13 must be parallel to the right vertical of Detail 2, the large rectangle, and it must be shifted to the left within Detail 13.

Detail 16: The horizontal line within Detail 13, which is a continuation of Detail 4 to the right, must come from the midpoint of the right side of the large rectangle and extend to the top of triangle 13. If triangle 13 is slightly askew, or if Detail 4 does not meet the midpoint of the right side of the rectangle, Detail 16 should still be scored as a full 2 points if it went to the top of the triangle from the midpoint of the right side of the rectangle.

Detail 17: The cross attached to the lower center area of the rectangle. The right side of the cross must be clearly longer than the left side of the cross but must not extend beyond the right end of the large rectangle. It should also, at its left end, commence at the midpoint of the right side of the square, which is Detail 18.

Detail 18: The square on the lower left corner of Detail 2. It must clearly be a square, as opposed to the rectangular shape of Detail 6, and its sides should be the same size as the vertical aspect of Detail 6, extending halfway between the left side of the rectangle and the vertical midline of the rectangle.

## Instruction Sheet For Dichotic Listening Test

The examinee is given a set of stereo headphones and showed the left ear headphone (marked with a “L”) and the right ear headphone. The volume for both ears is adjusted individually according to the subject’s preference so that the messages seem equally loud to the examinee. Verbal instructions are given about the contents of the test and what the examinee is supposed to do. Since a digit is most likely to be missed if it is presented when the examinee is speaking in response to the just preceding relevant digit, the examinee is advised to speak the relevant digit quickly as soon as he hears it so that he is ready for the next one. The examinee is told that he has to repeat aloud the “Test number” of the test so that the examiner knows which test is being played. Also, the examinee has to speak aloud only the digits from the relevant ear in both parts of the test and the examiner has to write these digits/letters. The relevant ear is designated by the tones (low tone=left ear is relevant; high tone=right ear is relevant). There are 4 practice tests. Before the practice tests, the examinee is played a “low” tone and a “high” tone so that he becomes familiar with it and can differentiate between them. Also, he is played a recording of the actual pronunciation of the digits (0 to 9) and alphabet letters (A to Z) so that he becomes familiar with their pronunciation and knows their peculiarities.

Out of these 4 practice tests, switching of the relevant channel is involved in two tests and in two tests the same ear is to be attended to in part one and part two. In the Dichotic Listening Test, it is standard practice to repeat the practice trials until the examinee knows what to do and the task has been understood. After the examinee understands the task, the actual tests start. There are 24 tests divided in 2 groups. Group one consists of “Test 1” to “Test 12”, and group two consists of “Test 13” to “Test 24”. After 12 tests, there is a break of 30 seconds and then the other 12 tests are instituted. To facilitate ambiguous-free scoring, a scoring sheet was developed which is shown in Table A.1.

**Table A.1** Dichotic Listening Test scoring sheet

| Dichotic Listening Test Scoring Sheet | | | | | | |
|---|---|---|---|---|---|---|
| | | Part I | Part I | Part II | Part II | Errors |
| Date | | | | | | |
| Purpose | | | | | | |
| ID | | | | | | |
| Name | | | | | | |
| Date of birth | | | | | | |
| occupation | | | | | | |
| | 4 7 6 0 2 8 - 5 9 3 | | | | | |
| | 5 1 9 3 - 4 7 6 | | | | | |
| * P Test 1 (LR) | | | | | | |
| | 2 4 6 3 7 0 - 1 9 5 | | | | | |
| | 9 8 - 7 4 2 | | | | | |
| P Test 2 (LL) | | | | | | |
| | 0 5 9 4 7 1 - 6 8 2 | | | | | |
| | 3 6 8 2 - 1 5 9 | | | | | |
| * P Test 3 (RL) | | | | | | |
| | 7 9 3 1 2 5 - 6 4 8 | | | | | |
| | 8 0 - 1 3 9 | | | | | |
| P Test 4 (RR) | | | | | | |
| | 3 4 6 8 2 0 - 9 7 1 | | | | | |
| | 1 9 7 5 - 3 8 4 | | | | | |
| Test 1 (RR) | | | | | | |

| Dichotic Listening Test Scoring Sheet | | | | | | |
|---|---|---|---|---|---|---|
| | | Part I | Part I | Part II | Part II | Errors |
| | 8 0 1 2 3 7 - 4 6 5 | | | | | |
| | 5 6 9 4 - 3 1 8 | | | | | |
| *Test 2 (LR) | | | | | | |
| | 0 4 1 9 3 5 - 6 2 8 | | | | | |
| | 8 7 - 3 0 5 | | | | | |
| *Test 3 (RL) | | | | | | |
| | 8 3 0 1 7 2 - 9 4 5 | | | | | |
| | 5 4 6 9 - 8 2 0 | | | | | |
| Test 4 (LL) | | | | | | |
| | 6 9 2 0 3 5 - 4 8 7 | | | | | |
| | 1 8 7 4 - 2 0 9 | | | | | |
| *Test 5 (LR) | | | | | | |
| | 6 2 7 0 9 8 - 5 3 4 | | | | | |
| | 3 5 1 4 - 2 8 7 | | | | | |
| Test 6 (RR) | | | | | | |
| | 5 3 9 0 1 6 - 2 4 8 | | | | | |
| | 7 4 2 8 - 1 9 6 | | | | | |
| Test 7 (RR) | | | | | | |
| | 6 5 7 2 4 1 - 9 0 3 | | | | | |
| | 3 8 - 2 5 6 | | | | | |
| Test 8 (LL) | | | | | | |
| | 7 9 6 1 5 3 - 2 0 8 | | | | | |
| | 4 0 2 8 - 3 1 9 | | | | | |
| Test 9 (LL) | | | | | | |
| | 8 5 9 6 3 0 - 7 4 2 | | | | | |

| Dichotic Listening Test Scoring Sheet | | | | | | |
|---|---|---|---|---|---|---|
| | | Part I | Part I | Part II | Part II | Errors |
| | 7 1 - 9 8 3 | | | | | |
| *Test 10 (RL) | | | | | | |
| | 4 8 5 9 1 0 - 7 2 3 | | | | | |
| | 6 3 - 8 4 5 | | | | | |
| Test 11 (RR) | | | | | | |
| | 6 9 7 2 3 0 - 1 5 8 | | | | | |
| | 5 1 4 8 - 0 7 2 | | | | | |
| Test 12 (LL) | | | | | | |
| | 4 2 0 5 1 8 - 3 6 7 | | | | | |
| | 6 9 - 1 0 2 | | | | | |
| *Test 13 (LR) | | | | | | |
| | 1 3 9 6 0 4 - 7 8 5 | | | | | |
| | 8 7 2 5 - 4 1 6 | | | | | |
| *Test 14 (LR) | | | | | | |
| | 4 3 1 6 8 5 - 2 0 9 | | | | | |
| | 9 7 - 5 4 1 | | | | | |
| Test 15 (LL) | | | | | | |
| | 9 4 2 5 6 8 - 0 7 3 | | | | | |
| | 3 1 - 9 5 8 | | | | | |
| Test 16 (LL) | | | | | | |
| | 5 0 8 3 1 9 - 4 6 7 | | | | | |
| | 4 2 - 1 5 9 | | | | | |
| *Test 17 (RL) | | | | | | |
| | 4 5 8 3 6 0 - 7 1 2 | | | | | |
| | 9 2 - 4 3 6 | | | | | |

| Dichotic Listening Test Scoring Sheet | | | | | | |
|---|---|---|---|---|---|---|
| | | Part I | Part I | Part II | Part II | Errors |
| Test 18 (RR) | | | | | | |
| | 4 2 8 3 0 9 - 1 5 6 | | | | | |
| | 7 6 - 3 8 4 | | | | | |
| *Test 19 (LR) | | | | | | |
| | 1 8 0 3 6 2 - 7 9 4 | | | | | |
| | 9 5 - 1 0 3 | | | | | |
| *Test 20 (LR) | | | | | | |
| | 2 6 7 1 4 0 - 9 8 5 | | | | | |
| | 8 3 - 1 2 4 | | | | | |
| Test 21 (RR) | | | | | | |
| | 7 1 4 8 3 9 - 6 2 0 | | | | | |
| | 0 6 5 2 - 3 9 7 | | | | | |
| *Test 22 (RL) | | | | | | |
| | 9 4 8 6 1 0 - 7 5 2 | | | | | |
| | 5 3 7 2 - 4 1 0 | | | | | |
| *Test 23 (RL) | | | | | | |
| | 3 4 0 2 6 7 - 8 9 5 | | | | | |
| | 8 1 9 5 - 0 4 7 | | | | | |
| *Test 24 (RL) | | | | | | |
| Total | | | | | | |

Note: The "P" before the Test No. stands for practice test.

* switching from left to right ear or from right to left ear takes place (row shown in shading).

## Scoring detail of the Dichotic Listening Test

To facilitate scoring of the Dichotic Listening Test, a score sheet was prepared (Table A.1). In this score sheet all the relevant digits for part one and part two of each test were written in bold font in one line. The irrelevant digits were noted in a line above the line of relevant digits. Tests wherein the switching of the relevant channel (from left to right or right to left) took place were given a shade to distinguish them from the rest, as switching errors are only noted in these tests (these tests were twelve in number). The examinee-dictated digits were noted on this sheet of paper in front of each test. To facilitate ambiguity-free scoring, we bifurcated the examinee-dictated digits into two groups (i.e., part one and part two of the test) by examining the actual relevant and irrelevant digits. For Omission errors, we noted the number of relevant digits in part one that were omitted by the examinees in each test and then added them up for all the tests; these were the Omission errors. We divided this by 72 and multiplied it by 100 to get percentage Omitted errors. We divided by 72 because in the 24 tests, 12 tests had 2 relevant digits and 12 had 4 relevant digits in part one.

For Intrusion errors, we inspected the examinee-dictated digits/letters in part one and compared them with the irrelevant digits/letters in part one. The number of common digits/letters between these comparisons was the intrusions. These were added for all 24 tests to obtain Intrusion errors. We divided this by 144 and multiplied it by 100 to get percentage Intrusion errors. We divided by 144 because in the 24 tests, each test had six irrelevant digits in part one, so the maximum number of intrusions that one could make was six in each test. For the switching errors category, since 50 percent of the second parts of the tests (i.e., 12 tests out of 24) involved switching to the other channel (the other ear), thus the number of Omission and Intrusion errors (as explained above) following the switch in relevant channel in the second part of these 12 tests were coded as switching errors. For these errors we had to consider only the 12 tests where the switching in the channel took place (i.e., Test number 2,3,5,10,13,14,17,19,20,22,23 and 24). The switching errors for these tests were added to get switching errors. We divided these by 72 and multiplied it by 100 to get percentage Switching errors. We divided by 72 because in the 12 tests under consideration, in each test there were 3 relevant digits in part two that could be missed and there were also three irrelevant digits that could have intruded.

To get the total number of errors we added the number of Omission, Intrusion and Switching errors for all 24 tests. Percentage Total errors were calculated by dividing this number by 288 (i.e., 72+144+72 = 288).

## Paper Folding Test Instructions

These instructions and the sample problem presented below have been adopted from Ekstrom et al. (1976) (Ekstrom, R.B., French, J.W., Harman, H.H. (1976). Manual for Kit of Factor-Referenced Cognitive Tests 1976. Educational Testing Service, Princeton, New Jersey.) In this test, it has to be imagined that pieces of paper are folded and then unfolded. Every problem in this test consists of some figures drawn to the left of a vertical line and some are drawn to the right of this vertical line. The figures on the left hand side of the vertical line show a square piece of paper being folded; the last of these figures has one or two holes drawn on it which shows the location where the paper has been punched. The hole is punched in such a manner that it passes through the whole of the thickness of the paper. There are five figures on the right hand side of the vertical line. One of these figures shows the position of the holes, when the paper is completely unfolded. You are to decide which of these five figures is the correct representation of the paper, when it is unfolded; the correct figure has to be marked with an X through it. The example in Figure A.5 and A.6 illustrates the problem.

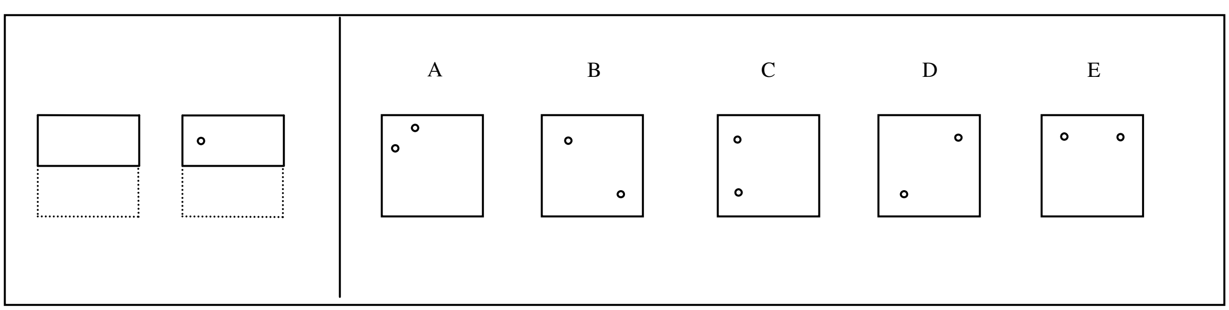



**Figure A.5** A simple example of Paper Folding test.

C is the correct answer to the sample problem presented in Figure A.5. Figure A.6 show the folding and unfolding of the paper and justifies the answer C, which is the correct answer.

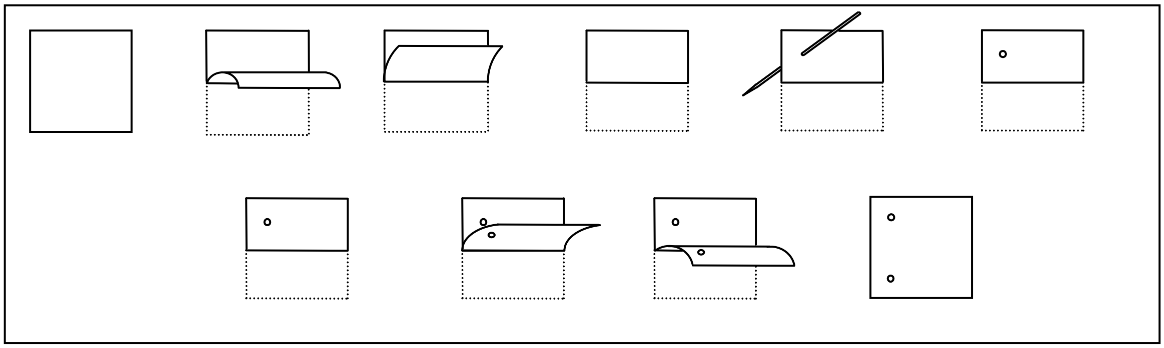


**Figure A.6** Showing unfolding of paper after punching of hole.

In all paper folding problems that follow, all the folds that the paper under goes are shown in the figures to the left of the vertical line and the paper is not moved or turned in any other way except to illustrate the folds shown in the figures. The answer is the figure that shows the position of the holes when the paper is completely unfolded.

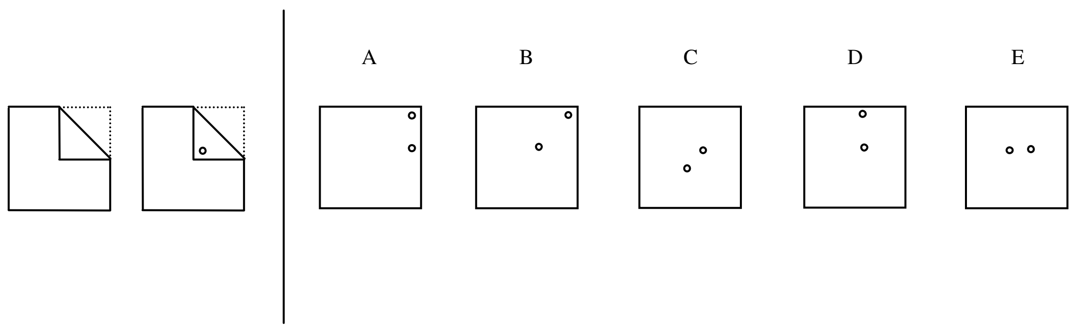
Scoring on the test will be such that the marks for correctly answered problems will add to the total score, while a fraction of the marks of each problem will be subtracted from the total score for each incorrect problem. Therefore, guessing will not be to the advantage of candidates unless they eliminate one or more of the answer alternatives as wrong. Also, paper folding problems that involve more folds carry more marks and vice versa .You will have 10 minutes to complete the test.

## Scoring of Paper Folding test

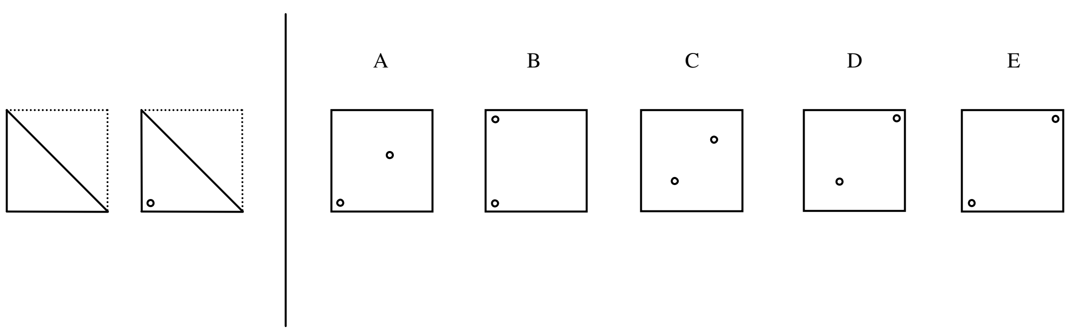
The examinees have 5 alternatives (i.e., A,B,C,D,E) to choose from on each test item. The 16 test items are not scored equally. The score of each test item (if correct) is proportional to the number of folds in that item. One fold item is allocated 2.5 marks (if correct), the two folds 5 marks, the 3 folds 7.5 marks and the 4 folds 10 marks. Since there are 4 items each of the one, two, three and four folds, if an examinee answers all of them correctly he gets 100 marks (i.e., $2.5 \times 4 + 5 \times 4 + 7.5 \times 4 + 10 \times 4 = 100$). To discourage guessing, some marks are subtracted from the total score, for incorrect answers. Since each item has 5 alternatives, the probability that a guess will be correct is one fifth; therefore if an item is incorrect, one fifth of the marks of that item are subtracted from the total score. For example, if an examinee answers all items correctly except one 3 fold item and one four fold item his net score is 79 (i.e., correct marks – incorrect marks = $(2.5 \times 4 + 5 \times 4 + 7.5 \times 3 + 10 \times 3) - (7.5 \times 1/5 + 10 \times 1/5) = 82.5 - 3.5 = 79$).

# Paper Folding Test

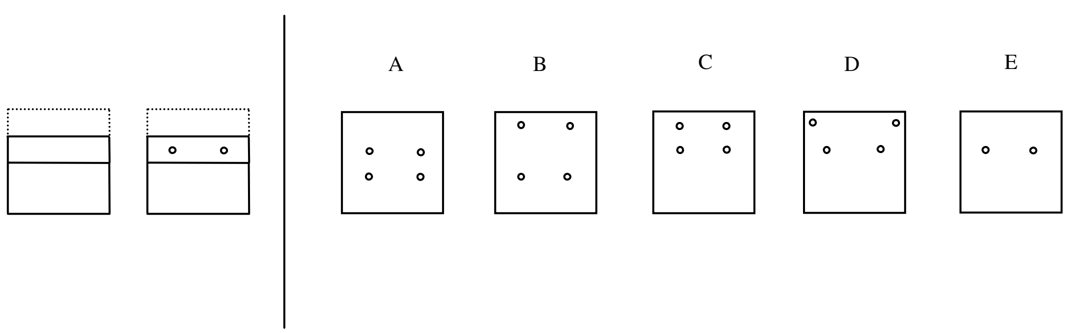
1



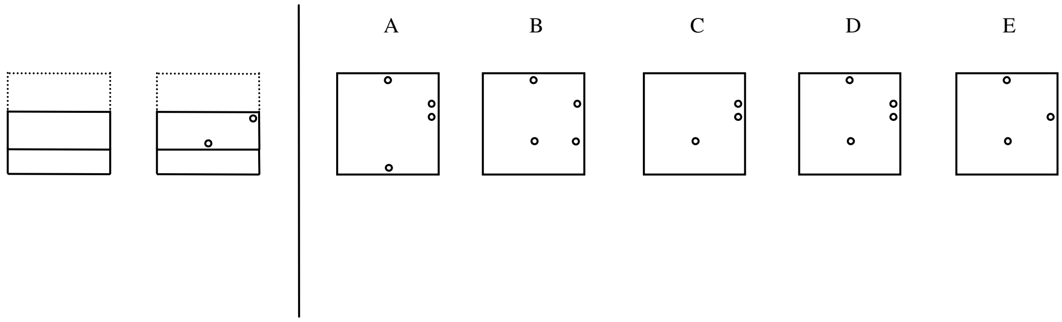

2



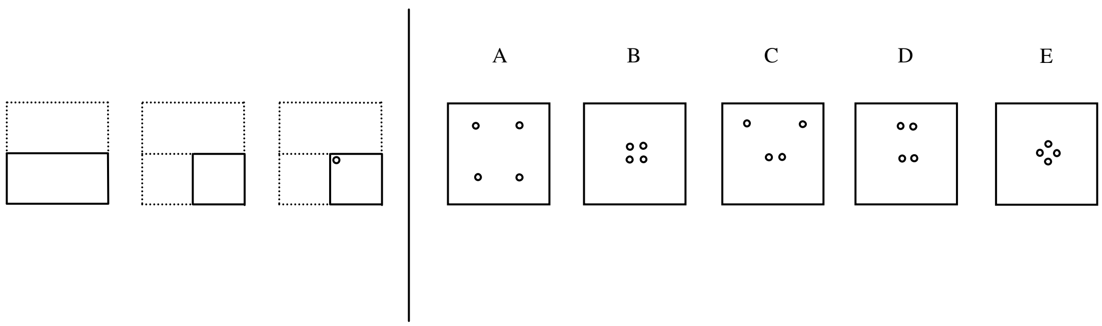

3



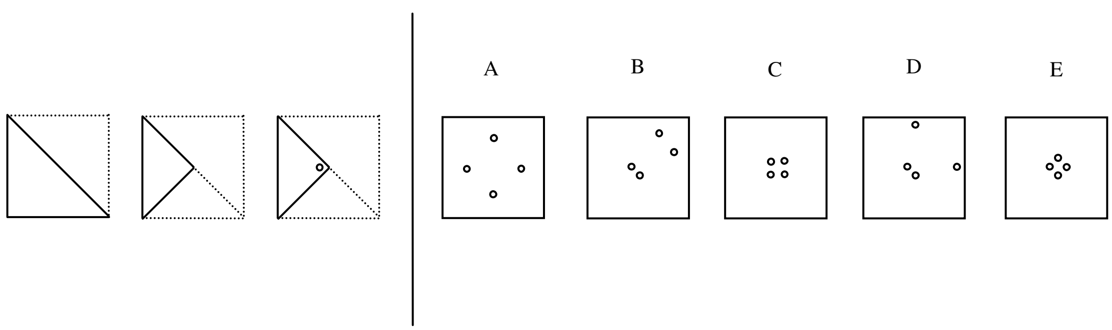
4



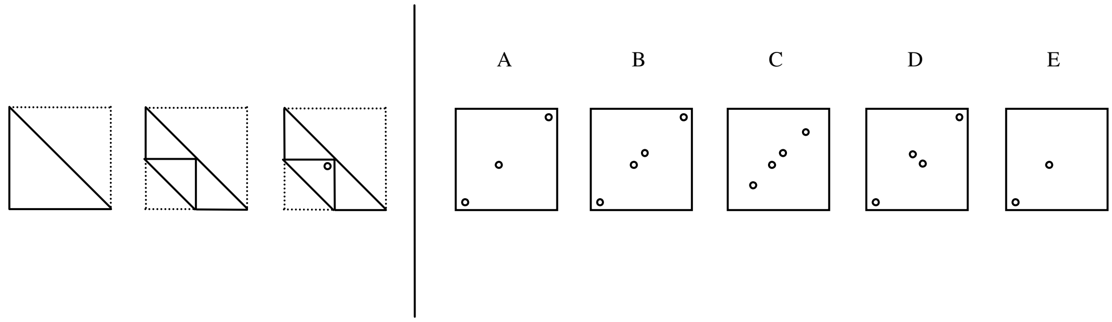

5



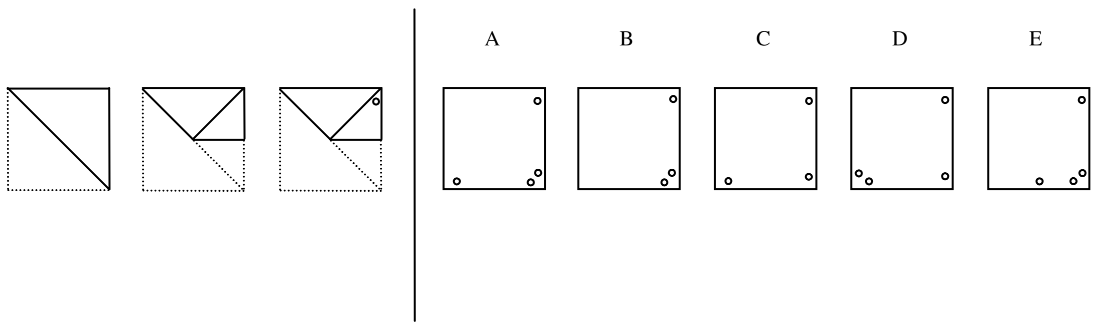

6



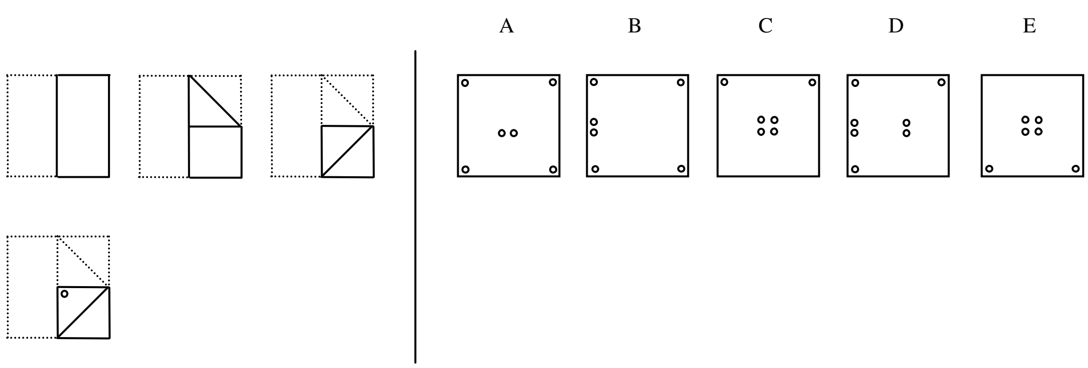
7



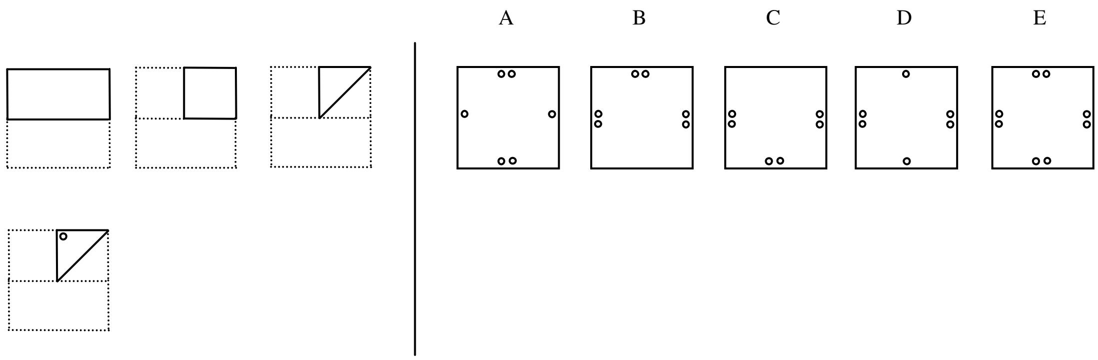

8



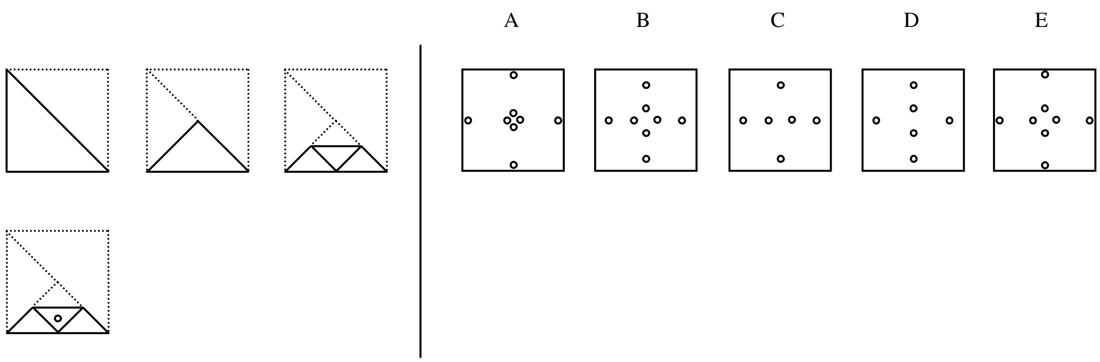

9



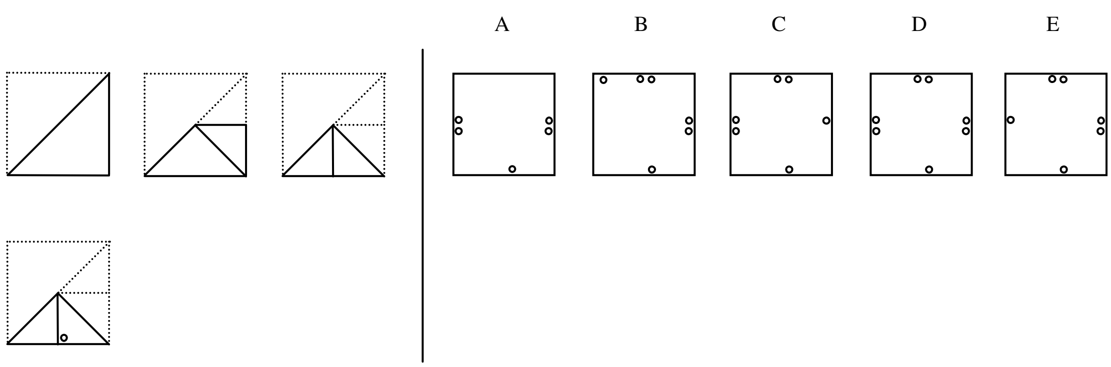
10



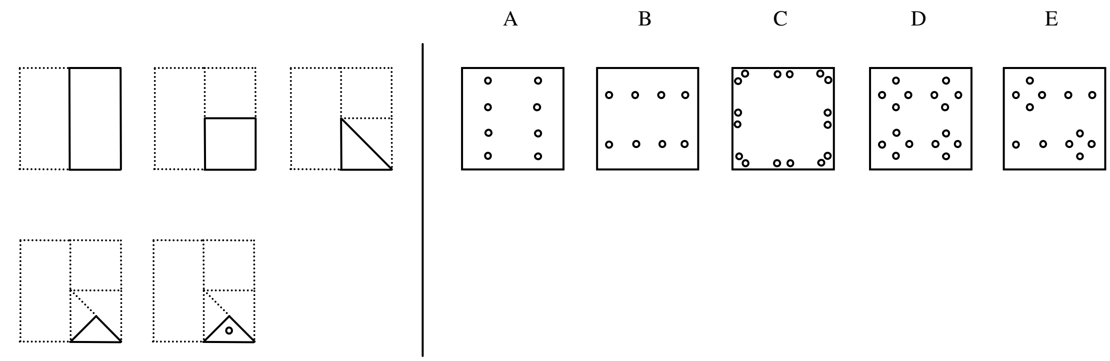

11



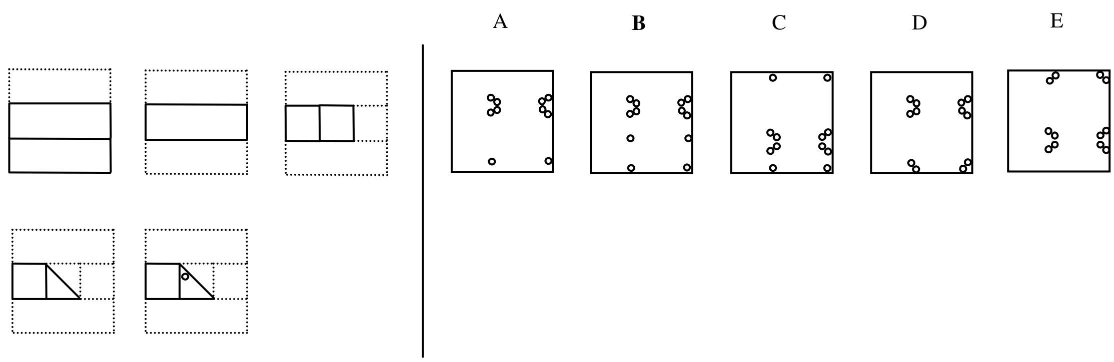

12



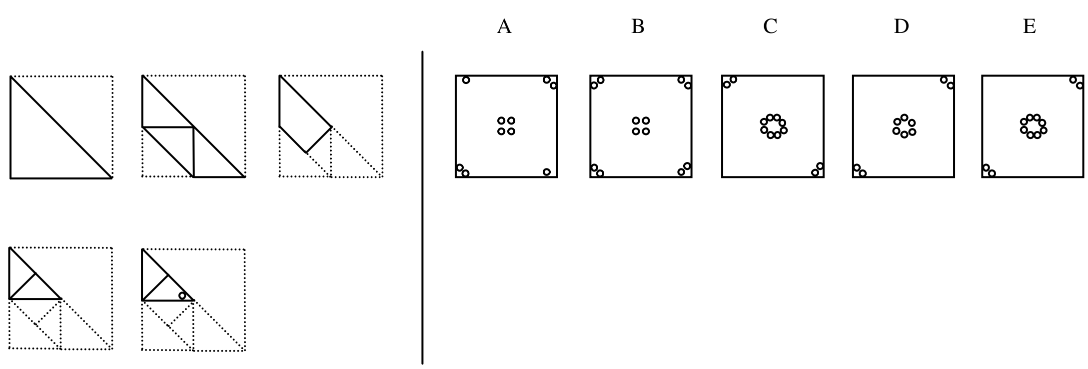
13



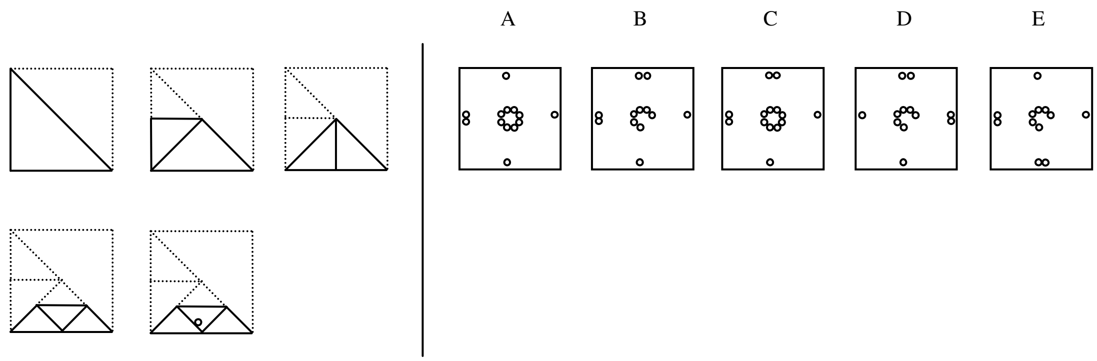

14




15



16




## UFOV Test Instructions

Presently, only the PC versions of the test are being marketed. We used the PC version that employs the mouse to make responses. For detailed testing procedure, the UFOV manual may be consulted (Visual Awareness Inc., 2002). However, before the test begins, examinees are informed that the test would not go forward unless a response is made, therefore if they are unsure of the correct answer, they should guess. UFOV is not a reaction time test. Therefore, an examinee should take his or her time in responding. Rather, it is the accuracy of his or her response that counts. Also, during the course of the test, there will be a point when the test becomes impossible and therefore, they should not be alarmed if they cannot recognize the displayed presentation at that time. Test scores for each subtset are generated by the computer. For interpretation of the scores, the UFOV manual may be consulted (Visual Awareness Inc., 2002). Table A.2 shows conversion of percent reduction in UFOV score (used in the Standard version) to subtest threshold in milliseconds (used in the PC versions) and vice versa (source: Edwards et al., 2005).

**Table A.2** Table for conversion of percent reduction in UFOV score (used in the Standard version) to subtest threshold in milliseconds (used in the PC versions) and vice versa (source: Edwards et al., 2005).

| UFOV Task | Standard Version (% reduction) | Touch and Mouse Versions (ms) |
|---|---|---|
| Subtest 1 | 0 | 0- 30 |
| | 5 | 31- 40 |
| | 10 | 41- 50 |
| | 15 | 51- 60 |
| | 20 | 61- 70 |
| | 25 | 71- 80 |
| | 30 | > 80 |
| Subtest 2 | 0 | < 20 |
| | 5 | 21- 40 |
| | 10 | 41- 80 |
| | 15 | 81-120 |
| | 20 | 121-160 |
| | 25 | 161-200 |
| | 27.5 | 201-240 |
| | 30 | > 240 |
| Subtest 3 | 0 | < 40 |
| | 7.5 | 41- 80 |
| | 12.5 | 81-120 |
| | 17.5 | 121-160 |
| | 22.5 | 161-200 |
| | 25 | 201-240 |
| | 30 | > 240 |
| Note: Total loss is the sum of the percent reduction from each of the three subtests | | |

.

# Appendix-B

Appendix-B pertains to the **Questionnaire**. Appendix-B is on Compact Disc

# Appendix-C

Appendix-C pertains to Chapter 4 i.e. **Driving Simulation**. Appendix-C is on Compact Disc

# Appendix-D

Appendix-D pertains to Chapter 5 i.e. **Driving Performance Indices**. Appendix-D is on Compact Disc

# Appendix-E

Appendix-E pertains to Chapter 6 i.e. **Cluster Analysis**. Appendix-E is on Compact Disc

# Appendix-F

Appendix-F pertains to Chapter 7 i.e. **Development of Linear Models**. Appendix-F is on Compact Disc

# Appendix-G

Appendix-G pertains to Chapter 8 i.e. **Logistic Regression Models**. Appendix-G is on Compact Disc

# References

Acock, A.C. (2006). *A Gentle Introduction to Stata*. Stata Press.

Adkins, G., Mercier, C., Mercier, J. and Shelley, M. (1999). Age and Gender as Predictors of Injury Severity in Broadside and Angle Vehicular Collisions. In *Transportation Research Record* 1693, TRB, National Research Council, Washington, D.C. 1999, pp.50-61.

Afifi, A., Clark, V.A. and May, S. (2003). *Computer-Aided Multivariate Analysis*. 4th Edition. Chapman & Hall/CRC.

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. 2nd Edition. Wiley.

Allen, R.W., Park, G., Rosenthal, T.J., and Aponso, B.L. (2004a). A process for Developing Scenarios for Driving Simulations. *IMAGE 2004 Conference*, Scottsdale, Arizona, July 2004.

Allen, R.W., Cook, M.L., Aponso, B.L., Rosenthal, T.J. (2004b). Special Issue Review of a Symposium on New Directions in Driving Simulation Research. *Advances in Transportation Studies* an International Journal 2004, special Issue: 5-8.

Allen, R.W., Marcotte, T.D., Rosenthal, T.J., and Aponso, B.L. (2005). Driver Assessment with Measures of Continuous Control Behaviour. *Third International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. Rockport, Maine, June 27-30, 2005.

Allen, M.J. and Yen, W.M. (1979). *Introduction to Measurement Theory*. Brooks/Cole Publishing Company.

Allison, P.D. (1995). *Survival Analysis Using the SAS® System: A Practical Guide*. SAS Institute Inc., Cary, NC, USA.

Alzheimer's Association. (2001). Driving. Available at: http://www.alz.org/advocacy/downloads/statements_driving.pdf.

Altman, D.G. (1991). *Practical Statistics for Medical Research*. Chapman & Hall.

Altman, D.G. and Royston, P. (2000). What do we mean by validating a prognostic model. *Statistics in Medicine*; 19: 453-473.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders*. 4th Edition. (DSM-IV).Washington D.C: APA.

American Heart Association. (2002). *Heart and Stroke Statistical Update*. Dallas, Texas.

Anstey, K., Hofer, S. and Luszcz, M. (2003). A Latent Growth Curve Analysis of late-Life Sensory and Cognitive Function over 8 Years: Evidence for Specific and Common Factors Underlying Change. *Psychology and Aging*; 18(4): 714-716.

Anderson, S.W., Rizzo, M., Shi, Q., Uc, E.Y. and Dawson, J.D. (2005). Cognitive Abilities Related to Driving Performance in a Simulator and Crashing on the Road. Driving Assessment 2005: *3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. Rockport, Maine, 27-30 June.

Anderson, S.W. and Tranel, D. (1989). Awareness of Disease States Following Cerebral Infarction, Dementia, and Head Trauma. *Clin Neuropsychol*; 3:327-339.

Avolio, B.J., Kroeck, K.G. and Panek, P.E. (1985). Individual Differences in Information-Processing Ability as a Predictor of Motor Vehicle Accidents. *Human Factors*; 27(5): 577-587.

Baddeley, A. (1986). *Working Memory*. New York: Oxford University Press.

Baddeley, A.D. (2000). The Magic Number and the Episodic Buffer. *Behavioral and Brain Sciences*; 24:117-118.

Baddeley, A.D.,Baddeley, H.A., Bucks, R.S., & Wilcock and G.K. (2001). Attention Control in Alzheimer's Disease. *Brain*; 124:1492-1508.

Bahill, A.T., Adler, D. and Stark, L. (1975). Most naturally Occurring Human Saccades have Magnitudes of 15 degrees or less. *Investigative Ophthalmology*; 14(6): 468-469.

Ball, K. and Owsley, C. (1993). The Useful Field of View Test: A New Technique for Evaluating Age-related Declines in Visual Function. *Journal of the American Optometric Association*; 64(1): 71-79.

Ball, K.K., Roenker, D.L., Wadley, V.G., Edwards, J.D., Roth, D.L., McGwin, G., Raleigh, R., Joyce, J.J., Cissell, G.M. and Dube, T. (2006). Can High-Risk Older Drivers Be Identified Through Performance-Based Measures In a department of Motor Vehicle Setting? *Journal of American Geriatrics Society*; 54: 77-84.

Ball, K., Owsley, C., Stalvey, B., Roenker, D.L., Sloane, M.E. and Graves, M. (1998). Driving Avoidance and Functional Impairment in Older Drivers. *Accident Analysis and Prevention*; 30(3): 313-322.

Ball, K., Owsley, C., Sloane, M.E., Roenker, D.L. and Bruni, J.R. (1993). Visual Attention Problems as a Predictor of Vehicle Crashes in Older Drivers. *Investigative Ophthalmology & Visual Science*; 34(11): 3110-3123.

Ball, K. and Owsley, C. (1991). Identifying Correlates of Accident Involvement for the Older Driver. *Human Factors*; 33: 583-595.

Ball K., Owsley, C. and Beard, B. (1990). Clinical Visual Perimetry underestimates peripheral field problems in older adults. *Clinical Vision Sciences*; 5: 113-125.

Ball, K.K., Vance, D.E., Edwards, J.D. and Wadley, V.G. (2004). Aging And The Brain. In M. Rizzo & P. J. Eslinger (Eds.), *Principles and Practice of Behavioral Neurology and Neuropsychology* (pp.795-809). Saunders-An imprint of Elsevier.

Ballard, J.L., Bishu, R.R., McCoy, P.T. and Tarawneh, M.S. (1993). Factors Associated with Driving Performance of Older Drivers. *In Transportation Research Record* 1405, TRB, National Research Council, Washington, D.C., pp.64-71.

Banfield, J. and Raftery, A.E. (1993). Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics*; 49: 803-821.

Barnes, M., Eames, P., Evans, C. et al.(1998): Rehabilitation After Traumatic Brain Injury: A Working Party Report of the British Society of Rehabilitation Medicine (London: BSRM).

Bautista, D., Arana, E., Marti-Bonmati, L. and Paredes, R. (1999). Validation of Logistic Regression Models in Small Samples: Applications to Calvarial Lesions Diagnosis. *Journal of Clinical Epidemiology*; 52(3): 237-241.

Belle, G.V. (2002). *Statistical Rules of Thumb*. Wiley.

Bennett, D.A., Wilson, R.S., Schneider, J.A., Evans, D.A., Beckett, L.A., Aggarwal, N.T. et al. (2002). Natural History of Mild Cognitive Impairment in Older Persons. *Neurology*; 59(2): 198-205.

Bergen, J.R., and Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*; 303(5919): 696-698.

Berge, J.M.F.T. and Hofstee, K.B. (1999). Coefficients Alpha and Reliabilities of Unrotated and Rotated Components. *Psychometrica*; 64(1): 83-90.

Berk, R.A. (2004). *Regression Analysis: A Constructive Critique*. Sage Publications.

Berlin, C.I. (1977). Hemispheric Asymmetry in Auditory Tasks. In S. Harnad, R.W. Doty, L. Goldstein, J. Jaynes, and G. Krauthamer (Eds.), *Lateralization in the Nervous System*. Academic Press, New York.

Berry, D.T.R. and Carpenter, G.S. (1992). Effect of Four Different delay periods on recall of the Rey-Osterrieth Complex figure by older persons. *The Clinical Neuropsychologist*; 6: 80-84.

Bieliauskas, L.A., Roper, B.R., Trobe, J., Green, P. and Lacy, M. (1998). Cognitive Measures, Driving Safety, and Alzheimer's Disease. *The Clinical Neuropsychologist*; 12: 206-212.

Bieliauskas, L.A., (2005). Neuropsychological Assessment of Geriatric Driving Competence. *Brain Injury*; 19(3): 221-226.

Birren, J.E. and Fisher, L.M. (1995). Aging and Speed of Behaviour: Possible Consequences for Psychological Functioning. *Annual Review of Psychology*; 46: 329-353.

Blajenkova, O., Kozhevnikov, M. and Motes, M.A.(2006). Object-Spatial Imagery: A New Self-Report Imagery Questionnaire. *Applied Cognitive Psychology*; 20: 239-263.

Bland, J.M. and Altman, D.G. (1997). Statistics Notes: Cronbach's Alpha. *British Medical Journal*; 314: 572

Boer, L.C., Harsveld, M. and Hermans, P.H. (1997). The Selective-Listening Task as a Test for Pilots and Air Traffic Controllers. *Military Psychology*; 9(2): 137-149.

Box, G.E.P. and Cox, D.R. (1964). An Analysis of Transformations. *Journal of Royal Statistical Society, Ser. B.*; 26: 211-246.

Box, G.E.P and Tidwell, P.W. (1962). Transformation of the Independent Variables. *Technometrics*; 4: 531-550.

Bozoki, A., Giordani, B., Heidebrink, J.L., Berent, S. and Foster, N.L. (2001). Mild Cognitive Impairments Predict Dementia in Non-demented Elderly Patients with Memory Loss. *Arch Neurol.*; 58(3): 411-416.

British Psychological Society. (2001). *Fitness to Drive and Cognition.* British Psychological Society. St. Andrews House, 48 Princes Road East Leicester LE1 7DR, UK.

Brodaty, H. and Moore, C.M. (1997). The Clock Drawing Test for Dementia of the Alzheimer's Type: A Comparison of Three Scoring Methods in a Memory Disorders Clinic. *International Journal of Geriatric Psychiatry*; 12: 619-627.

Brookhuis, K. and Waard, D.D. (1994). Measuring Driving Performance by Car-Following in Traffic. *Ergonomics*; 37(3): 427-434.

Brown, I.D., Tickner, A.H. and Simmonds, D.C.V. (1969). Interference between concurrent tasks of driving and telephoning. *J Appl Psychol*; 53: 419.

Brown, L.B. and Ott, B.R. (2004). Driving and Dementia: A Review of the Literature. *Journal of Geriatric Psychiatry and Neurology*; 17: 232-240.

Brown, L.B., Ott, B.R., Papadonatos, G.D., et al. (2005). Prediction of On-road Driving Performance in Patients with Early Alzheimer's Disease. *Journal of American Geriatrics Society*; 53: 94-98.

Brouwer, W.H. and Withaar, F.K. (1997). Fitness to Drive After Traumatic Brain Injury. *Neuropsychological Rehabilitation*; 7(3): 177-193.

Byington, S., D. Harkey, K. Lococo, and L. Staplin. (2001) *Guidelines and Recommendations to Accommodate Older Drivers and Pedestrians*. FHWA-RD-01-051. FHWA, U.S. Department of Transportation.

Caird, J.K., Edwards, C. J., Creasser, J. I. and Horrey, W. J. (2005). Older Driver Failures of Attention at Intersections: Using Change Blindness Methods to Access Turn Decision Accuracy. *Human Factors*; 47(2): 235-249.

Caird, J. K. and Hancock, P.A. (2002). Left Turn and gap acceptance accidents. In R.E. Dewar & P. L. Olson (Eds.), *Human Factors in Traffic Safety* (pp.613-652). Tucson, AZ:Lawyers & Judges.

Cairney, P. and Catchpole, J. (1996). Patterns of Perceptual Failures at Intersections of Arterial Roads and Local Streets. In A.G. Gale, I.D. Brown, C.M. Haslegrave, & S.P. Taylor (Eds.), *Vision In Vehicles –V* (pp.87-94). Amsterdam: Elsevier Science.

Carr, D.B. (1997). Motor Vehicle Crashes and Drivers with DAT. *Alzheimer Disease and Associated Disorders*; 11(1): 42-47.

Carr, D., Jackson, T. and Alquire, P. (1990). Characteristics of an Elderly Driver Population Referred to a Geriatric Assessment Center. *Journal of American Geriatrics Society*; 38: 1145-1150.

Carr, D.B., Duchek, J.M., Meuser, T.M. and Morris, J.C. (2006). Older Adult Drivers With Cognitive Impairment. *American Family Physician*; 73(6): 1029-1034.

Carroll, J. (1993). *Human cognitive abilities: A survey of factor-analytic Studies*. New York: Cambridge University Press.

Carter, J.E., Obler, L., Woodward, S. and Albert, M. (1983). Effect of Increasing Age on the Latency of Saccadic Eye Movements. *Journal of Gerontology*; 38: 318-320.

Cavallo, V., Laya, O., and Laurent, M.(1986). The Estimation of Time-to-Collision as a function of Visual Stimulation. In: A.G. Gale et al.(eds.) *Vision in Vehicles*, Elsevier Science Publishers BV: Amsterdam, Netherlands.

Celeux, G. and Govaert, G. (1993). Comparison of the Mixture and the Classification Maximum Likelihood in Cluster Analysis. *Journal of Statistical Computation and Simulation*; 47: 127-146.

Chandraratna, S. and Stamadiadis, N. (2003). Problem driving Maneuvers of Elderly Drivers. *Transportation Research Record* 1843: 89-95.

Chaparro, A. and Alton, J. (2000). Age Related Differences in Driving Performance and Target Identification. *Proceedings of IEA 2000 / HFES 2000 Congress*, San Diego, CA; 4: 56-59.

Chatterjee, S. and Hadi, A.S. (2006). *Regression Analysis by Example*. 4th Edition. Wiley-Interscience.

Chervinsky, A., Mitrushina, M. and Satz, P. (1992). Comparison of four methods of scoring the Rey-Osterrieth Complex figure drawing test on four age groups of normal elderly. *Brain Dysfunction*; 5: 267-287.

Chovan, J.D., Tijerina, L., Everson, J.H., Pierowicz, J.A. and Hendricks, D.L. (1994). *Examination of Intersection Left Turn Across Path Crashes and IVHS countermeasures* (DOT HS 808 154). Cambridge, MA: National Highway Traffic Safety Administration.

Christensen, R. (2001). *Advanced Linear Modeling*. 2nd Edition. Springer.

Christie, N. (1996). Assessing Driving Fitness Following Brain Injury or Illness: A Research review. *TRL Report 208*: 1-21.

Christie, N., Savill, T., Grayson, G. et al. (2001a). The Assessment of Fitness to Drive after Brain Injury or Illness. Department of Environment, *Transport and Regions, Transport Research Laboratory Report no. 485*, ISSN 0968-4107.

Christie, N., Savill, T., Buttress, S., Newby, G. and Tyerman, A. (2001b). Assessing Fitness to Drive After Head Injury: A Survey of Clinical Psychologists. *Neuropsychological Rehabilitation*; 11(1): 45-55.

Clay, O.J., Wadley, V.G., Edwards, J.D., Roth, D.L., Roenker, D.L. and Ball, K.K. (2005). Cumulative Meta-analysis of the Relationship between Useful Field of View and Driving Performance in Older Adults: Current and Future Implications. *Optometry and Vision Science*; 82(8): 724-731.

Cohen, A., Gnanadesikan, R., Kettenring, J.R. and Landwehr, J.M. (1977). Methodological Developments in Some Applications of Clustering, in *Applications of Statistics* (P.R. Krishnaiah, ed.). North-Holland, Amsterdam.

Cohen, J., Cohen, P., West, S.G., and Aiken, L.S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd Edition. Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

Cook, R.D. and Weisberg, S. (1999). *Applied Regression including Computing and Graphics*. Wiley.

Cooper, P.J. (1990). Differences in Accident Characteristics Among Elderly and Middle-Aged Drivers. *Accident Analysis and Prevention*; 22: 499-508.

Cox, D.J., Quillian, W.C., Thorndike, F.P., Kovatchev, B.P. and Hanna, G. (1998). Evaluating Driving Performance of Outpatients with Alzheimer's Disease. *Journal of the American Board of Family Practice*; 11(4): 264-271.

Craik, F.I.M. and Rabinowitz, J.C. (1984). Age Differences in the Acquisition and use of verbal information: A tutorial review. In H. Buouma & D.G. Bouwhuis (Eds.), *Attention and Performance X* (pp.471-499). Hillsdale, N.J: Erlbaum.

Craik, F.I.M. and Salthouse, T.A. (1992). *The Handbook of Aging and Cognition*. Lawrence Erlbaum, Publishers. Hillsdale, New Jersey.

Crawford, J.R., Parker, D.M. and McKinlay, W.W. (1992). *A Handbook of Neuropsychological Assessment*. Lawrence Erlbaum Associates, Publishers. Hillsdale (USA).

Cummings, J.L. and Cole, G. (2002). Alzheimer's disease. *Journal of the American Medical Association*; 287: 2335-2358.

Daigneault, G., Joly, P. and Frigon, J. (2002). Executive Function in the the Evaluation of Accident Risk of Older Drivers. *Journal of Clinical and Experimental Neuropsychology*; 24: 221-238.

Davey, G. (1987). *Cognitive Processes and Pavlovian Conditioning in Humans*. John Wiley & Sons.

Davis, H. and Rockwood, K. (2004). Conceptualization of mild cognitive impairment: A Review. *International Journal of Geriatric Psychiatry*; 19: 313-319.

Davies, D.R. and Parasuraman, R. (1982). *The Psychology of Vigilance*. London, Academic Press.

Dawson, L.K. and Grant, I. (2000). Alcoholics' initial organizational and problem-solving skills predict learning and memory performance on the Rey-Osterrieth Complex figure. *Journal of the International Neuropsychological Society*; 6: 12-19.

DeLong, E. R., DeLong, D.M. and Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating curves: A nonparametric approach. *Biometrics*; 44: 837-845.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum Likelihood for Incomplete Data via the EM Algorithm. *Journal of Royal Statistical Society B*; 39: 1-38.

Department of Transport (DOT). (1997). *Road Accident Statistics Great Britain: The Casualty Report*. HMSO, London.

Department of Transport (DOT). (2001). *Older Drivers: A literature review (No.25)*. HMSO, London.

De Raedt, R. and Ponjaert-Kristoffersen, I. (2000). The Relationship Between Cognitive/Neuropsychological Factors and Car Driving Performance in Older Adults. *Journal of American Geriatrics Society*; 48: 1664-1668.

De Raedt, R. and Ponjaert-Kristoffersen, I. (2001). Predicting at-fault Car Accidents of Older Drivers. *Accident Analysis and Prevention*; 33: 809-819.

DeVellis, R.F. (2003). *Scale Development: Theory and Applications*. 2nd Edition. Sage Publications.

Dikmen, S.S., Heaton, R.K., Grant, I. and Temkin, N.R. (1999). Test-retest Reliability and Practice Effects of Expanded Halstead-Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society*; 5: 346-356.

Dobbs, A. R., Robert, B.H. and Schopflocher, D. (1998). A comparative Approach to Identify Unsafe Older Drivers. *Accident Analysis and Prevention*; 30(3): 363-370.

Drachman, D.A., Swearer, J.M., the Collaborative Group. (1993). Driving and Alzheimer's Disease: The Risk of Crashes. *Neurology*; 43: 2448-2456.

Drakopoulos, A. and Lyles, R.W., (1997). Driver Age as a Factor in Comprehension of Left-Turn Signals. *Transportation Research Record* No. 1573: 76-85.

Draper, N.R. and Smith, H. (1998). *Applied Regression Analysis*. 3rd Edition. John Wiley & Sons.

Dubinsky, R.M., Stein, A.C. and Lyons, K. (2000). Practice Parameter: Risk of Driving and Alzheimer's Disease (an evidence-based review): *Report of the Quality Standards Subcommittee of the American Academy of Neurology*; 54: 2205-2211.

Duchek, J.M., Hunt, L., Ball, K., Buckles, V. and Morris, J.C. (1998). Attention and Driving Performance in Alzheimer's Disease. *Journal of Gerontology: Psychological Sciences*; 53B (2): 130-141.

Duchek, J.M. and Balota, D.A. (2005). Failure to Control Prepotent Pathways in Early Stage Dementia of the Alzheimer's Type: Evidence from Dichotic Listening. *Neuropsychology*; 19(5): 687-695.

Dulisse, B. (1997). Older Drivers and Risk to Other Road Users. *Accident Analysis and Prevention*; 29: 573-582.

Edmans, J.A. and Lincoln, N.B. (1987). Frequency of Perceptual Deficits After Stroke. *Clinical Rehabilitation*; 1: 273-281.

Edwards, J.D., Vance, D.E., Wadley, V.G., Cissell, G.M., Roenker, D.L. and Ball, K.K. (2005). Reliability and Validity of Useful Field of View Test Scores as Administered by Personal Computer. *Journal of Clinical and Experimental Neuropsychology*; 27: 529-543.

Efron, B. (1986). How Biased is the apparent error rate of a prediction rule. *Journal of American Statistical Association*; 81: 461-470.

Ekstrom, R.B., French, J.W. and Harman, H.H. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests* 1976. Educational Testing Service, Princeton, New Jersey.

Esteban-Santillan,C., Praditsuwan, R., Ueda, H. and Geldmacher, D. (1998). Clock Drawing Test in Very Mild Alzheimer's Disease. *Journal of American Geriatrics Society*; 46(10): 1266-1269.

Evans, L. (1991). *Traffic Safety and the Driver*. New York: Van Nostrand Reinhold.

Everitt, B.S. (1996). An Introduction to Finite Mixture Distributions. *Statistical Methods in Medical Research*; 5: 107-127.

Everitt, B.S. (2005). *An R and S-PLUS Companion to Multivariate Analysis*. Springer.

Everitt, B.S., Landau, S. and Leese, M. (2001). *Cluster Analysis*. Arnold . London.

Everitt, B.S. and Dunn, G. (2001). *Applied Multivariate Data Analysis*. 2nd Edition. Arnold. London.

Everitt, B.S. and Dunn, G. (1991). *Applied Multivariate Data Analysis*. Edward Arnold, London.

Everitt, B.S. and Hothorn, T. (2006). *A Handbook of Statistical Analyses using R*. Chapman & Hall/CRC.

Fals-Stewart, W. (1991). *An Interrater Reliability Study of the Trail Making Test (Part A and B)*. Unpublished manuscript.

Faraway, J.J. (2005). *Linear Models with R*. Chapman & Hall/CRC.

Ferman, T.J., Boeve, B.F., Smith, G.E., et al. (1999). REM Sleep Behavior Disorder and Dementia: Cognitive Differences when Compared With AD. *Neurology*; 52: 951-957.

Ferrucci, L., Cecchi, F., Guralnik, J.M., Giampaoli, S., Noce, C.L., Salani, B., Bandinelli, S. and Baroni, A. (1996). Does the Clock Drawing Test Predict Cognitive Decline in Older Persons Independent of the Mini-Mental State Examination? *Journal of American Geriatrics Society*; 44(11): 1326-1331.

Fellgiebel, A., Wille, P., Muller, M., Winterer, G., Scheurich, A., Vucurevic, G., Schmidt, L. and Stoeter, P. (2004). Ultrastructural Hippocampal and White Matter Alterations in Mild Cognitive Impairement: A diffusion Tensor Imaging Study. *Dementia & Geriatric Cognitive Disorders*; 18(1): 101-108.

Fiorentino, D.D. and Parseghian, Z. (1997). Time-To-Collision: A Sensitive Measure of Driver Interaction With Traffic in a Simulated Driving Task. Paper 553. *Human Factors and Ergonomics Society 41st Annual Meeting*, Albuquerque, New Mexico.

Fisher, D.L., Laurie, N.E., Glaser, R., Connerney, K., Pollatsek, A., Duffy, S.A. and Brock, J. (2002). Use of a Fixed-base Driving Simulator to Evaluate the Effects of Experience and PC-based Risk Awareness Training on Driver's Decisions. *Human Factors*; 44: 287-302.

Fisk, G.D., Schneider, J.J. and Novack, T.A. (1998). Driving Following Traumatic Brain Injury: prevalence , Exposure, Advice and Evaluation. *Brain Injury*; 12: 683-695.

Fisk, G.D., Novack, T., Mennemeier, M. and Roenker, D. (2002). Useful Field of View after Traumatic Brain Injury. *Journal of Head Trauma and Rehabilitation*; 17(1): 16-25.

Fitten, L.J., Perryman, K.M., Wilkinson, C., et al. (1995). Alzheimer and Vascular Dementias and Driving: A Prospective Road and Laboratory Study. *JAMA*; 273:1360-1365.

Fitten, L.J. (2003). Driver Screening for Older Adults. *Archives of Internal Medicine*; 163(18): 2129-2132.

Fletcher, R. H and Fletcher, S.W. (2005). *Clinical Epidemiology: The Essentials*. 4th Edition. Lippincott Williams & Wilkins.

Foos, P.W. and Wright, L. (1992). Adult Age Differences in the Storage of Information in Working Memory. *Exp Aging Res*; 2: 51-57.

Fox, G.K., Bowden, S.C., Bashford, G.M. and Smith, D.S. (1997). Alzheimer's Disease and Driving: Prediction and Assessment of Driving Performance. *Journal of American Geriatrics Society*; 45: 949-953.

Fox, G.K., Withaar, F. and Bashford, G.M. (1996). Dementia and Driving: A Survey of clinical Practice in Aged Care Assessment Teams. *Australian Journal on Ageing*; 15: 111-114.

Fox, G., Bowden, S. and Smith, D. (1998). On-road Assessment of Driving Competence After Brain Impairment: Review of Current Practice and Recommendations for Standardized Examination. *Archives of Physical Medicine and Rehabilitation*; 79: 1288-1296.

Fozard, J.L., Vercruyssen, M., Reynolds, S.L., Hancock, P.A. and Quilter, R.E. (1994). Age Differences and Changes in Reaction Time: The Baltimore Longitudinal Study of Aging. *Journal of Gerontology: Psychological Sciences*; 49: 179-189.

Fraley, C. (1998). Algorithms for Model-Based Gaussian Hierarchical Clustering. *SIAM Journal of Scientific Computing*; 20: 270-281.

Fraley, C. and Raftery, A.E. (1998). How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. *The Computer Journal*; 41(8): 578-588.

Fraley, C. and Raftery, A.E. (1999). MCLUST: Software for Model-Based Cluster Analysis. *Journal of Classification*; 16: 297-306.

Fraley, C. and Raftery, A.E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association*; 97(458): 611-631.

Fraley, C. and Raftery, A.E. (2006). *MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering. Technical Report No. 504*. Department of Statistics, University of Washington. Revised (minor) November 2007.

Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K.I. and Delis, D.C. (1994). *Clock Drawing: A Neuropsychological Analysis*. Oxford University Press.

Freedman, D.A. (2005). *Statistical Models: Theory and Practice*. Cambridge University Press.

Freund, B., Gravenstein, S. and Ferris, R. (2002). Evaluating Driving Performance of Cognitively Impaired and Healthy Older Adults: A Pilot Study Comparing On-road Testing and Driving Simulation. *Journal of American Geriatrics Society*; 50: 1309.

Freund, B., Gravenstein, S., Ferris, R., Burke, B.L. and Shaheen, E. (2005a). Drawing Clocks and Driving Cars. *Journal of General Internal Medicine*; 20(3): 240-244.

Freund, B., Colgrove, L.A., Burke, B.L. and Mcleod, R. (2005b). Self-rated Driving Performance Among Elderly Drivers Referred for Driving Evaluation. *Accident Analysis and Prevention*; 37: 613-618.

Friedman, H.P. and Rubin, J. (1967). On Some Invariant Criteria for Grouping Data. *Journal of the American Statistical Association*; 62: 1159-1178.

Fujii, D.E., Lloyd, H.A. and Miyamoto, K. (2000). The salience of Visuospatial and organizational skills in reproducing the Rey-Osterrieth Complex Figure in subjects with high and low IQ. *The Clinical Neuropsychologist*; 14: 551-554.

Funahashi, S. (2001). Neuronal Mechanisms of Executive Control by the Prefrontal Cortex. *Neuroscience Research*; 39:147-165.

George, C.F.P. (2003). Driving Simulators in Clinical Practice. *Sleep Medicine Reviews*; 7: 311-320.

Gilfillan, C.K. and Schwartzberg, J.G. (2005). Addressing the At-Risk Older Driver. *Clinical Geriatrics*; 13(8): 27-34.

Goode, K.T., Ball, K.K., Sloane, M., Roenker, D.L., Roth, D.L., Myers, R.S. and Owsley, C. (1998). Useful Field of View and Other Neurocognitive Indicators of Crash Risk in Older Adults. *Journal of Clinical Psychology in Medical Settings*; 5(4): 425-440.

Gopher, D., and Kahneman, D. (1971). Individual Differences in Attention and the Prediction of Flight Criteria. *Perceptual and Motor Skills*; 33: 1335-1342.

Gopher, D. (1982). A Selective Attention Test as a Predictor of Success in Flight Training. *Human Factors*; 24(2): 173-183.

Gordon, A.D. (1999). *Classification*. Chapman & Hall/CRC.

Grady CL, Haxby JV, Horwitz B, Sundaram M, Berg G, Schapiro M, et al. (1988). Longitudinal study of the early neuropsychological and cerebral metabolic changes in dementia of the Alzheimer type. *J Clin Exp Neuropsychol*; 10: 576–96.

Graham, J.W., Hofer, S.M. and Piccinin, A.M. (1994). Analysis with missing data in drug prevention research. In Collins, L.& Seitz, L. (Eds.), *National Institute on Drug Abuse Research Monograph Series, Vol. 142*, pp. 13–62. National Institute on Drug Abuse, Washington, D.C.

Green, S.B. (1991). How Many Subjects Does it Take to Do a Regression Analysis? *Multivariate Behavioral Research*; 26:499-510.

Greenwood, P.M., Parasuraman, R. and Alexander, G.E. (1997). Controlling the Focus of Spatial Attention During Visual Search: Effects of Advanced Aging and Alzheimer Disease. *Neuropsychology*; 11(1): 3-12.

Groeger, J.A. (2000). *Understanding Driving: Applying Cognitive Psychology to a Complex Everyday Task*. Psychology Press. Taylor & Francis Group.

Groth-Marnat, G. (2000). *Neuropsychological Assessment in Clinical Practice*. John Wiley & Sons, Inc.

Guerrier, J.H., Manivannan, P. and Nair, S.N. (1999). The Role of Working Memory, Field Dependence, Visual Search, and Reaction Time in the Left Turn Performance of Older Female Drivers. *Applied Ergonomics*; 30: 109-119.

Hakamies-Blomqvist, L. (1996). Research on Older Drivers: A Review. *Journal of International Association of Traffic and Safety Sciences*; 20: 91-101.

Hakamies-Blomqvis, Wilkund, M. and Henriksson, P. (2005). Predicting older drivers'accident involvement – Smeed's Law revisited. *Accident Analysis and Prevention*; 37: 675-680.

Hakamies-Blomqvist, L. (1993). Fatal Accidents of Older Drivers. *Accident Analysis and Prevention*; 25(1): 19-27.

Hakamies-Blomqvis, L. (1994). Compensation in Older Drivers as Reflected in their Fatal accidents. *Accident Analysis and Prevention*; 26(1): 107-112.

Hakamies-Blomqvist, L. (2004). Safety of Older Persons in Traffic. Transportation in an Aging Society. *Conference Proceedings 27. Transportation Resaerch Board*. pp.22-35.

Hamby, S.L., Wilkins, J.W. and Barry, N.S. (1993). Organizational Quality on Rey-Osterrieth and Taylor Complex Figure tests: A New Scoring system. *Psychological Assessment*; 5(1): 27-33.

Hamilton, L.C. (2005). *Statistics with Stata*. Duxbury Resource Center.

Hartigan, J.A. and Wong, M.A. (1979). A K-Means Clustering Algorithm. *Applied Statistics*; 28: 100-108.

Hartman, E. (1970). *Driver Vision Requirements. SAE Technical Paper Series, 700392*.

Hartman, M. and Potter, G. (1998). Sources of Age Differences on the Rey-Osterrieth Complex Figure Test. *The Clinical Neuropsychologist*; 12: 513-524.

Hebert, L.E., Scherr, P.A., Bienias, J.L., Bennett, D.A. and Evans, D.A. (2003). Alzheimer Disease in the US Population: Prevalence Estimates using the 2000 Census. *Archives in Neurology*; 60(8): 1119-1122.

Hegarty, M., Shah, P. and Miyake, A. (2000). Constraints on using the Dual-task Methodology to Specify the Degree of Central Executive Involvement in Cognitive Tasks. *Memory & Cognition*; 28(3): 376-385.

Heinik, J., Solomesh, I., Shein, V. and Becker, D. (2002a). Clock Drawing Test in Mild and Moderate Dementia of the Alzheimer's type: a comprehensive and correlation study. *International Journal of Geriatric Psychiatry*; 17: 480-485.

Heinik, J., Solomesh, I., Raikher, B. and Lin, R. (2002b).Can the clock drawing test help to differentiate between dementia of the Alzheimer's type and vascular dementia? A preliminary study. *International Journal of Geriatric Psychiatry*; 17: 699-703.

Heinik, J., Solomesh, I. and Berkman, P. (2004). Correlation between the CAMCOG, the MMSE, and Three Clock Drawing Tests in a Specialized outpatient psychogeriatric service. *Archives of Gerontology and Geriatrics*; 38: 77-84.

Helmstadter, G.C. (1964). *Principles of Psychological Measurement*. Prentice Hall.

Herrmann, N., Kidron, D., Shulman, K., Kaplan, E., Binns, M., Soni, J., Leach, L. and Freedman, M. (1999). The use of the Clock Tests in Schizophrenia. *General Hospital Psychiatry*; 21: 70-73.

Heyman, A., Fillenbaum, G.G., Welsh-Bohmer, K.A., et al. for the Consortium to Establish a Registry for Alzheimer's Disease. (1998). Cerebral infarcts in Patients with autopsy-proven Alzheimer's disease. *Neurology*; 51(pt 18): 159-162.

Hills, B.L. and Johnson, L. (1980). *Speed and Minimum Gap Acceptance Judgements at Two Rural Junctions. Report SR515*. Department of the Environment, Transport and Road Research Laboratory: Crowthorne, Berkshire, England
.
Hopewell, C.A. (2002). Neuropsychological, Medical and Psychiatric Disorders Affecting Motor Vehicle Operations. In R.E. Dewar & P. L. Olson (Eds.), *Human Factors in Traffic Safety* (pp.235-289). Tucson, AZ:Lawyers & Judges.

Horst, P. (1966). *Psychological Measurement and Prediction*. Wadsworth Publishing Company. Belmont, California.

Hosmer, D.W. and Lemeshow, S. (2000). *Applied logistic Regression*. 2nd Edition. John Wiley & Sons, Inc.

Houwelingen, J.C.V and Cessie, S.L. (1990). Predictive Value of Statistical Models. *Statistics in Medicine*; 9: 1303-1325.

Hu, P.S., Trumble, D., and Lu, A. (1995). *Driving Decisions and Vehicle Crashes Among Older Drivers*. Oak Ridge National Laboratory, Oak Ridge.

Huang, C. (2003). *Mild Cognitive Impairment: Neuroimaging Markers for Early Diagnosis of Dementia*. Karolinska Institute, Department of Clinical Neuroscience, Occupational Therapy and Elderly Care Research (NEUROTEC), Division of Geriatric Medicine, Huddinge University Hospital Stockholm.

Hunt, L.A., Morris, J.C., Edwards, D. and Wilson, B.S. (1993). Driving Performance in Persons With Mild Senile Dementia of the Alzheimer Type. *Journal of American Geriatrics Society*; 41: 747-753.

Hunt, L.A. (1994). Driving and the Demented Person. In John C. Morris (ed). *Handbook Of Dementing Illnesses*. Marcel Dekker, Inc.

Hunt, L.A., Murphy, C.F., Carr, D., Duchek, J.M., Buckles, V. and Morris, J.C. (1997a). Reliability of the Washington University Road Test. *Archives of Neurology*; 54(6): 707-712.

Hunt, L.A., Murphy, C.F., Carr, D., Duchek, J.M., Buckles, V., and Morris, J.C. (1997b). Reliability of the Washington University Road Test: A Performance-Based Assessment for Drivers with Dementia of the Alzheimer Type. *Archives of Neurology*; 54: 707-712.

Hunt, L.A., Murphy, C.F., Carr, D., Duchek, J.M., Buckles, V., and Morris, J.C. (1997c). Environmental Cueing May Affect Performance on a Road Test for Drivers with Dementia of the Alzheimer Type. *Alzheimer Disease and Associated Disorders*; 11 (suppl. 1): 13-16.

Inouye, S.K., Albert, M.S., Mohs, R.C., Sun, K. and Berkman, L.F. (1993). Cognitive Performance in a high-functioning Community-Dwelling Elderly Population. *J Gerontol A Biol Sci Med Sci.*; 48: M146-M151.

Insurance Institute for Highway Safety (IIHS) (1992) *Status Report Special Issue: Crashes, Fatal Crashes per Mile*. Insurance Institute of Highway Safety, Washington, DC.

Insurance Institute for Highway Safety. (2000). *IIHS Fatality Facts: Elderly*. Washington, DC: Author.

Isler, R.B., Parsonson, B.S. and Hansson, G.J. (1997). Age Related Effects of Restricted Head Movements on the Useful Field of View of Drivers. *Accident Analysis and Prevention*; 29(6): 793-801.

JAMA. (2003). Public Health and Aging-United States and Worldwide. *Journal of the American Medical Association*; 289 (11): 1371-1373.

James, W. (1890). *The Principles of Psychology. Vol.1*. New York : Holt; 403-404.

Janke, M.K. (1994). *Age Related Disabilities That May Impair Driving and their Assessment*. Report Under Cooperative Agreement DTNH22-93-Y-5330. U.S. Department of Transportation, Washington, DC.

Janke, M.K. and Eberhard, J.W. (1998). Assessing Medically Impaired Older Drivers in a Licensing Agency Setting. *Accident Analysis & Prevention*; 30(3): 347-361.

Jewell, N.P. (2004). *Statistics for Epidemiology*. Chapman & Hall/CRC.

Johansson, K. and Lundberg, C. (1997). The 1994 International Consensus Conference on Dementia and Driving: A brief report. *Alzheimer Disease and Associated Disorders*; 11: 62-69.

Johansson, K., Bogdanovic, N., Kalimo, H., Winblad, B. and Viitanen, M. (1997). Alzheimer's Disease and Apolipoprotein E Epsilon 4 Allele in Older Drivers who Died in Automobile Accidents. *Lancet*; 349: 1143-1144.

Johansson, E., and Engstrom, J., Cherri, C., Nodari, E., Toffetti , Schindhelm, R. and Gelau, C. (2004). *Review of Existing Techniques and Metrics for IVIS and ADAS Assessment.* AIDE IST-1-507674-IP . Information Society Technologies (IST).

Kahneman, D., Ben-Ishai, R. and Lotan, M. (1973). Relation of a Test of Attention to Road Accidents. *Journal of Applied Psychology*; 58(1): 113-115.

Kaplan, E. (1988). A Process Approach to Neuropsychological Assessment. In T. Bull & B.K. Bryant (Eds.), *Clinical Neuropsychology and Brain Function: Research, Measurement, and Practice* (pp.129-167). Washington, D.C.: American Psychological Association.

Kaplan, E. (1990). The Process Approach to Neuropsychological Assessment of Psychiatric Patients. *Journal of Neuropsychiatry*; 2: 72-87.

Kaplan, R.M. and Saccuzzo, D.P. (2001). *Psychological Testing: Principles, Applications, and Issues.* 5th Edition. Wadsworth.

Kass, R.E. and Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association*; 90(430): 773-795.

Kawas, C. (2003). Early Alzheimer's Disease. *New England Journal of Medicine*; 349: 1056-1063.

Kinsella  k. and Velkoff,  V. (2001). U.S. Census Bureau. An ageing World: 2001. Washington, D.C: U.S. Government Printing Office, 2001; series P95/01-1.

Kito, T., Haraguchi, M. and Funatsu, T. (1989). Measurements of Gaze Movements While Driving. *Perceptual and Motor Skills*; 68: 19-25.

Kline, P. (1993). *The Handbook of Psychological Testing*. Routledge, London and New York.

Kline, D.W., Kline, T. J.B., Fozard, J.L., Kosnik, W., Schieber, F. and Sekuler, R. (1992). Vision, Aging, and Driving: The Problems of Older Drivers. *Journal of Gerontology*; 47(1): 27-34.

Knight, J.A., Kaplan, E. and Ireland, L.D. (2003). Survey Findings of Rey-Osterrieth Complex Figure usage. In J.A. Knight (Ed.), *The Handbook of Rey-Osterrieth Complex Figure Usage: Clinical and Research application*. Lutz, Fla.: PAR.

Knoblauch, R.L., Nitzburg, M., Reinfurt, D., Council, F., Zegeer, C. and Popkin, C. (1995). *Traffic Operations Control for Older Drivers*, FHWA-RD-94-119. Federal Highway Administration, McLeans, VA.

Knopman, D.S. and Nissen, M.J. (1987). Implicit learning in patients with probable Alzheimer's Disease. *Neurology*; 37: 784-788.

Korteling, J.E. and Kaptein, M.A. (1996). Neuropsychological Driving Fitness Tests for Brain Damaged Subjects. *Archives of Physical Medicine and Rehabilitation*; 77: 138-146.

Kray J., Lindenberger, U., (2000). Adult age differences in Task Switching. *Psychology and Aging*; 15: 126-147.

Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W. (2005). Applied Linear Statistical Models. 5/e McGraw-Hill.

La Rue, A. (1992). *Critical Issues in Neuropsychology: Aging and neuropsychological Assessment.* New York: Plenum Press.

Lachman, R., Lachman, J. and Butterfield, E.C. (1979). *Cognitive Psychology and Information Processing: An Introduction*. Lawrence Erlbaum Associates, Publishers. Hillsdale, New Jersey.

Lambert, L.D. and Fleury, M. (1994). Age, Cognitive Style, and Traffic Signs. *Perceptual and Motor Skills*; 78(2): 611-624.

Lee, D.N. (1976). A Theory of Visual Control of Braking Based on Information about Time-to-Collision. *Perception*; 5: 437-459.

Lee, H.C., Drake, V. and Cameron, D. (2002a). Identification of Appropriate Assessment Criteria to Measure Older Adults' Driving Performance in Simulated Driving. *Australian Occupational Therapy Journal*; 49: 138-145.

Lee, H.C., Cameron, D. and Lee, A.H. (2002b). Assessing the Driving Performance of Older Adult Drivers: On-road versus Simulated Driving. *Accident Analysis and Prevention*; 894: 1-7.

Lee, H., Swanwick, R.J., Coen, R.F. and Lawlor, B.A. (1996). Use of the Clock Drawing Task in the Diagnosis of Mild and Very Mild Alzheimer's Disease. *Int. Psychogeriat.*; 8: 469-476.

Leese, M. and Landau, S. (2006). Model-based Clustering using S-Plus. *International Journal of Methods in Psychiatric Research*; 15(3): 146-156.

Leibowitz, H.W. (1986). Grade Crossing accidents and human factors engineering. *American Scientist*; 73: 558-562.

Lew, H.L., Poole, J.H., Lee, E.H., Jaffe, D.L., Huang, H. and Brodd, E. (2005). Predictive Validity of Driving-Simulator Assessments following Traumatic Brain Injury: A Preliminary Study. *Brain Injury*; 19(3): 177-188.

Lezak, M. (1983). *Neuropsychological Assessment*. 2nd Edition. Oxford: Oxford University Press.

Lezak, M., Howieson, D.B. and Loring, D.W. (2004). *Neuropsychological Assessment*. 4th Edition. Oxford: Oxford University Press.

Li, H., Rosenthal, R. and Rubin, D.B. (1996). Reliability of Measurement in Psychology: From Spearman-Brown to Maximal Reliability. *Psychological Methods*; 1(1): 98-107.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. John Wiley & Sons, New York.

Long, J.S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications.

Long, J.S and Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata*. 2nd Edition. Stata Press.

Lord, F.M. (1958). Some Relations Between Guttman's Principal Components of Scale Analysis and Other Psychometric Theory. *Psychometrika*; 23: 291-296.

Lucas-Blaustein, M.J., Filipp, L., Dungan, C. and Tune, L. (1988). Driving in Patients with Dementia. *Journal of the American Geriatrics Society*; 36: 1087-1091.

Lundberg, C., Hakamies-Blomqvist L., Almkvist, O. and Johansson, K., (1998) Impairments of Some Cognitive Functions are Ccommon in Crash-involved Older Drivers. *Accident Analysis and Prevention*; 30(3): 371-377.

Lundberg, C., Johansson, K., Ball, K., Bjerre, B., Blomqvist, C., Braekhus, A., Brouwer, W.H., Bylsma, F.W., Carr, D.B., Englund, L., Friedland, R.P., Hakamies-Blomqvist, L., Klemetz, G., O'Neill, D., Odenheimer, G.L., Rizzo, M., Schelin, M., Seideman, M., Tallman, K., Viitanen, M.,Waller, P.F. and Winblad, B. (1997). Dementia and Driving: An attempt at Consensus. *Alzheimer Disease and associated Disorders*, 11(1): 28-37.

Lundberg, C. (2003). *Older Drivers With Cognitive Impairments: Issues of Detection and Assessment*. From Neurotec Department, Division of Clinical Geriatrics, Karolinska Institute, Huddinge University Hospital, Stockholm, Sweden.

Lundqvist, A. (2001). Neuropsychological Aspects of Driving. *Brain Injury*; 15(11): 981-994.

Lundqvist, A., Alinder, J., Alm, H. et al. (1997). Neuropsychological Aspects of Driving After Brain Lesion: Simulator Study and on Road Driving. *Applied Neuropsychology*; 4: 220-230.

Lundqvist, A. and Ronnberg, J. (2001). Driving Problems and Adaptive Driving Behaviour After Brain Injury: A Qualitative Assessment; *Neuropsychological Rehabilitation*; 11:171-185.

Lutkepohl, H. (1985). Comparison of criteria for estimating the order of a vector autoregressive process. *Journal of Time Series Analysis*; 65: 297-303.

Lyman, H.B. (1998). *Test Scores and What They Mean*. Allyn and Bacon.

Mackay (1988). *Crash protection for older persons. Transportation in an ageing society, Special Report 218*. Transportation Research Board, National Research Council, Washington, DC, pp. 158-193.

Mallows, C. (1973). Some comments on $C_p$. *Technometrics*; 15: 661-676.

Manly, B.F.J. (2005). *Multivariate Statistical Methods: A Primer*. 3rd Edition. Chapman & Hall/CRC.

Marottoli, R.A., Cooney, L.M., Wagner, D.R. et al. (1994). Predictors of automobile crashes and moving violations among elderly drivers. *Ann Intern Med*; 121: 842-846.

Masur, D.M., Sliwinski, M., Lipton, R.B., Blau, A.D. and Crystal, H.A. (1994). Neuropsychological Prediction of Dementia and the Absence of Dementia in Healthy Elderly Persons. *Neurology*; 44: 1427-1432.

Matlin, M.W. (2005). *Cognition*. 6th Edition. John Wiley & Sons, Inc.

Matthew, G., Davies, D.R., Westerman, S.J. et al. (2000). *Human Performance. Cognition, Stress and Individual Differences*. Hove Psychology Press.

McDowell, I. (2006). *Measuring Health: A Guide to Rating Scales and Questionnaires*. 3rd Edition. Oxford University Press.

McDowd, J.M. and Filion, D.L. (1992). Ageing, Selective Attention , and Inhibitory Processes: A Psychological Approach. *Psychology and Aging*; 7: 65-71.

McDowd, J.M. and Craik, F.I.M. (1988). Effects of Aging and Task Difficulty on Divided Attention Performance. *J Exp Psychol Hum Percept Perf.*; 14: 267

McKenna, P. (1998). Fitness to Drive: A neuropsychological Perspective. *Journal of Mental Health*; 7: 9-18.

Mcknight, J.A. (1988). Driver and Pedestrian training. *Transportation in an Aging Society. Improving mobility and Safety for Older Persons, Vol. 2*, Technical Papers. Transportation Research Board, National Research council, Washington, D.C.

McDowd, J.M. and Shaw, R.J. (2000). Attention and Aging: A Functional Perspective. In F.I.M. Craik & T.A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (2nd ed., pp.221-292). Mahwah, NJ: Erlbaum.

McKenna, F.P., Duncan, J. and Brown, I.D. (1986). Cognitive Abilities and Safety on the Road: A Re-examination of Individual Differences in Dichotic Listening and Search for Embedded Figures. *Ergonomics*; 29(5): 649-663.

McKnight, A. J. and McKnight, A.S. (1999). Multivariate analysis of age-related driver ability and performance deficits. *Accident Analysis and Prevention*; 31: 445-454.

McKnight, A. J. and McKnight, A.S. (1993) The Effect of Cellular Phone Use Upon Driver Attention. *Accident Analysis and Prevention*; 25(3): 259-265.

Mcknight A.J. and Urquijo, J.I. (1993). Signs of Deficiency Among Elderly Drivers. *Transportation Reasearch Record* 1405: 81-86.

McGwin, G. and Brown, D.B. (1999). Characteristics of Traffic Crashes among Young, Middle-Aged, and Older Drivers. *Accident Analysis and Prevention*; 31: 181-198.

McLachlan, G.J. and Chang, S.U. (2004). Mixture Modelling for Cluster Analysis. *Statistical Methods in Medical Research*; 13: 347-361.

McLachlan, G.J and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley, New York.

Mendez, M.F., Ala, T. and Underwood, K.L. (1992). Development of Scoring Criteria for the Clock Drawing Task in Alzheimer's Disease. *Journal American Geriatrics Society*; 40: 1095-1099.

Meyers, J.E. and Meyers, K.R. (1995a). Rey Complex Figure Test under four different administration procedures. *The Clinical Neuropsychologist*; 9: 63-67.

Meyers, J.E. and Meyers, K.R. (1995b). *The Meyers Scoring System for the Rey Complex Figure and the Recognition trial: Professional Manual*. Odessa, Fla.: Psychological Assessment Resources.

Michon, J.A. (1985). A Critical View of Driver Behavior Models: What Do We Do, What Should We Do? In *Human Behavior and Traffic Safety* (L. Evans and R. Schwing, eds.), Plenum Press, New York, pp.485-520.

Mihal, W.L. and Barret, G.V. (1976). Individual Differences in Perceptual Information Processing and Their Relation to Automobile Accident Involvement. *Journal of Applied Psychology*; 61(2): 229-233.

Mirsky, A.F., Anthony, B.J., Duncan, C.C., Ahearn, M.B. and Kellam, S.G. (1991). Analysis of the elements of Attention: A Neuropsychological Approach. *Neuropsychol Rev*; 2: 109-145.

Mitrushina, M.M., Boone, K.B., Razani, J. and D'Elia L.F. (2005). *Handbook of Normative Data for Neuropsychological Assessment*. 2nd Edition. Newyork: Oxford University Press.

Morris, J.C. (1993). The Clinical Dementia Rating (CDR): Current Version and Scoring Rules. *Neurology*; 43: 2412-2414.

Morrison, J.H. and Hof, P.R. (1997). Life and Death of  Neurons in the Aging Brain. *Science*; 278: 412-419.

Morris, J.C., Storandt, M., Miller, J.P., Mckeel, D.W., Price, J.L., Rubin, E., H. and Berg, L. (2001). Mild Cognitive Impairment Represents Early-Stage Alzheimer Disease. *Arch Neurol.*; 58(3): 397-405.

Montgomery, D.C., Peck, E.A. and Vining, G.G. (2006). *Introduction to Linear Regression Analysis*. 4th Edition. Wiley-Interscience.

Moore, R. L., Sedgley, I. P. and Sabey, B.E., (1982). *Ages of Car Drivers Involved in Accidents, With Special Reference to Junctions.* Transport and Road Research Laboratory Supplementary Report 718. Transport Studies Group, University College, London, England.

Moore, V. and Wyke, M.A.(1984). Drawing Disability in Patients with Senile Dementia. *Psychol Med*; 14: 97-105.

Mourant, R.R. (1979). Driving Performance of the Elderly. *Accident Analysis and Prevention;* 11: 247-253.

Mourant, R.R. and Thattacherry, T.R. (2000). Simulator Sickness in a Virtual Environments Driving Simulator. *Proceedings of the IEA 2000/HFES 2000 Congress*.

Myers, R.S., Ball, K.K., Kalina, T.D., Roth, D.L. and Goode, K.T. (2000). Relation of Useful Field of View and Other Screening Tests to On-road Driving Performance. *Perceptual and Motor skills*; 91: 279-290.

Naftali, R. (2000). Aging of the Brain and its Impact on Cognitive Performance: Integration of Structural and Functional Findings. In F.I.M. Craik & T.A. Salthouse (Eds.), *The Handbook of Aging and Cognition*. (p.57). Lawrence Erlbaum Associates. London.

NHTSA. (1997). *Traffic Safety Facts 1996: Older Population*, National Highway Traffic Safety Administration, Washington, DC, 4 pp.

Nilsson, L.G. (2003). Memory Function in Normal Aging. *Acta Neurologica Scandinavica*; 107 (s179): 7-13.

Noble, B. (2000). *Travel characteristics of older people. Transport Trends*. DTLR.

Nouri, F.M., Tinson, D.J. and Lincoln, N.B. (1987). Cognitive Ability and Driving after Stroke. *Int. Disabil. Studies*; 9: 110-115.

Nunnally, J.C. (1978). *Psychometric Theory*. 2nd edition . New York: McGraw Hill.

Nunnally, J.C. and Bernstein, I.H. (1994). *Psychometric Theory*. 3rd Edition. McGraw-Hill.

O'Regan, J.K., Rensink, R.A. and Clark, J.J. (1999). Change Blindness as a result of "mudsplashes". *Nature*; 398: 34.

OECD (2001). *Ageing and Transport: mobility needs and safety issues*. Paris, OECD.

O'Neill, D., Neubauer, K., Boyle, M., Gerrard, J., Surmon, D. and Wilcock, G.K. (1992). Dementia and Driving. *Journal of the Royal Society of Medicine*; 85: 199-202.

OPCS. (1992). *1989 based National Population Projections*, PP2, NO17. HMSO.

Osterrieth, P.A. (1944). Le test de copie d'une figure complex: Contribution a l'etude de la perception et de la memoire. *Archives de Psychologie*; 30: 286-356.

Oxley, P.R. (1991). *Elderly and disabled car users: A growing market in private transport for the elderly. TRL CR308.*, Crowthorne, Berkshire.

Owsley, C. (2004). In: *Transportation in an Aging Society: A Decade of Experience. Conference Proceedings 27*. Transportation Research Board, Washington, D.C.

Owsley, C., Ball, K., Sloane, M.E., Roenker, D.L. and Bruni, J.R. (1991). Visual Cognitive Correlates of Vehicle Accidents in Older Drivers. *Psychology and Aging*; 6(3): 403-415.

Owsley, C., Ball, K. and Keeton, D.M. (1995). Relationship between visual sensitivity and target localization in older adults. *Vision Research*; 35(4): 579-587.

Owsley, C., Ball, K., McGwin, G., Sloane, M.E., Roenker, D.L., White, M.F., and Overley, E.T. (1998a). Visual Processing Impairment and Risk of Motor Vehicle Crash Among Older Adults. *The Journal of the American Medical Association (JAMA)*; 279 (14): 1083-1088.

Owsley, C., McGwin, G. and Ball, K. (1998b). Vision Impairment, Eye Disease, and Injurious Motor Vehicle Crashes in the Elderly. *Ophthalmic Epidemiology*; 5(2): 101-113.

Papacostas, C.S. and Prevedouros, P.D. (2005). *Transportation Engineering and Planning*. Prentice Hall.

Papageorgiou, I. and Baxter, M.J. (2001). Model-Based Cluster Analysis of Artifact Compositional Data. *Archaeometry*; 43(4): 571-588.

Parasuraman, R. and Nestor, P.G. (1991). Attention and Driving Skills in Alzheimer's Disease. *Human factors*; 33: 539-557.

Parasuraman, R. and Nestor, P.G. (1993). Attention and Driving: Assessment in Elderly Individuals with Dementia. *Clinics in Geriatric Medicine*; 9(2): 377-387.

Parasuraman, R. (1998). *Probing the Attentive Brain at Work and in Disease. Psychological Science Agenda (American Psychological Association)*; November/December: 6-8.

Parasuraman, R., Greenwood, P.M., Haxby, J. and Grady, C.L. (1992). Visuospatial Attention in Dementia of the Alzheimer Type. *Brain*; 115(3): 711-733.

Parasuraman, R., Greenwood, P.M. and Alexander, G.E. (2000).Alzheimer Disease constricts the dynamic range of Spatial Attention. *Neuropsychologia*; 38:1126-1135.

Park, G.D., Allen, R.W., Rosenthal, T.J. and Fiorentino, D. (2007). Older Driver Simulator Performance in Relation to Driving Habits and DMV Driving Records. *2nd International Conference on Technology and Aging*, Toronto, Canada, 2007.

Patomella, A. and Kottorp, A. (2005). An Evaluation of Driving Ability in a simulator: A good Predictor of Driving Ability after Stroke? *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Rockport, Maine.*

Perloff, J.M. and Persons, J.B. (1988). Biases Resulting From the Use of Indexes: An Application to Attributional Style and Depression. *Psychological Bulletin*; 103(1): 95-104.

Perry, R.J. and Hodges, J.R. (1999). Attention and Executive Deficits in Alzheimer's Disease: A Critical Review. *Brain*; 122: 383:404.

Petersen, R.C., Doody, R., Kurz, A., Mohs, R.C., Morris, J.C., Rabins, P.V., Ritchie, K., Rossor, M., Thal, L. and Winblad, B. (2001). Current Concepts in Mild Cognitive Impairment. Archives of Neurology; 58: 1985-1992.

Petersen, R.C., Smith, G.E., Ivnik, R.J., Kokmen, E. and Tangalos, E.G. (1994). Memory Function in Very Early Alzheimer's Disease. *Neurology*; 44: 867-872.

Petersen, R. (2003). *Mild Cognitive Impairment. Aging to Alzheimer's Disease*. New York: Oxford University Press.

Pidikiti, R.D. nad Novack, T.A. (1991). The Disabled Driver: An Unmet Challenge. *Archives of Physical Medicine and Rehabilitation*; 72: 109-111.

Ponds, R.W.H.M., Brouwer, W.H. and Van Wolffelaar, P.C. (1988). Age Differences in Divided Attention in a Simulated Driving Task. *Journal of Gerontology*; 43: 151-156.

Posner, M.I. (1980). Orienting of Attention. *Q J Exp Psychol*; 32: 3-25.

Powlishta, K.K., Dras, V.D.D., Stanford, A., Carr, D.B. and Tsering, C. (2002). The clock Drawing Test is a Poor Screen for Very Mild Dementia. *Neurology*; 59: 898-903.

Preusser, D. F., Williams, A.F., Fergusan, S. A., Ulmer, R.G. andWeinstein, H. B. (1998). Fatal Crash Risk For Older Drivers At Intersections. *Accident Analysis and Prevention*; 30(2): 151-159.

Price, J.L., Davis, P.B., Morris, J.C. and White, D.L. (1991). The Distribution of Tangles, Plaques and Related immunohistochemical Markers in Healthy Aging and Alzheimer's Disease. *Neurobiology of Aging*; 12: 295-312.

Quillian, W.C., Cox, D.J., Kovatchev, B.P. and Phillips, C. (1999). The Effects of Age and Alcohol Intoxication on Simulated Driving Performance, Awareness and Self-restraint. *Age and Aging*; 28: 59-66.

Radford, K.A., Lincoln, N.B. and Murray-Leslie, C. (2004). Validation of the Stoke Drivers Screening Assessment for People with Traumatic Brain Injury. *Brain Injury*; 18(8): 775-786.

Ranney, T.A. (1994). Models of Driving Behaviour: A Review of Their Evolution. *Accident Analysis & Prevention*; 26: 733-750.

Ratrout, N.T., Al-Ofi, K.A. and Iyaz, S. (2004). Validation and Improvement of a Rear-End Conflict Prediction Model. *Transportation Research Record: Journal of the Transportation Research Board, No. 1897*: pp. 206-210.

Ray, W.A., Thapa, P.B. and Shorr, R.I. (1993). Medications and the Older Driver. *Clin Geriatr Med*; 9: 413-438.

Reger, M.A., Welsh, R.J., Watson, G.S., Cholerton, B., Baker, L.D. and Craft, S. (2004). The Relationship Between Neuropsychological Functioning and Driving Ability in Dementia: A Meta-Analysis. *Neuropsychology*; 18(1): 85-93.

Reitan, R.M. (1958). Validity of Trail Making Test as an Indicator of Organic Brain Damage. *Perceptual and Motor Skills*; 8: 271-276.

Rencher, A.C. (2002). *Methods of Multivariate Analysis*. 2nd Edition. Wiley-Interscience.

Retchin,S.M. and Hillner, B.E. (1994). The Costs and Benefits of a Screening Program to Detect Dementia in Older Drivers. *Med Decis Making*; 14: 315-324.

Ritchie, K., Artero, S., and Touchon, J. (2001). Classification Criteria for Mild Cognitive Impairment. *Neurology*; 56: 37-42.

Rizzo, M., McGehee, D. V., Dawson, J.D. and Anderson, S. N. (2001). Simulated Car Crashes at Intersections in Drivers with Alzheimer Disease. *Alzheimer Disease and Associated Disorders*; 15(1): 10-20.

Rizzo, M. (2004). Safe and Unsafe Driving. In *Principles and Practice of Behavioral neurology and neuropsychology* (M. Rizzo and P.J. Eslinger, eds.), W.B. Saunders, Philadelphia, Pa., pp. 197-222.

Rizzo, M. and Kellison, I.L. (2004). Eyes, Brain, and Autos. *Archives of Ophthalmology*; 122: 641-647.

Rizzo, M., Reinach, S., McGehee, D. V. and Dawson, J.D. (1997). Simulated Car Crashes and Crash Predictors in Drivers with Alzheimer's Disease. *Arch Neurol.*; 54: 545-551.

Rizzo, M., Stierman, L., Skaar, N., Dawson, J.D., Anderson, S.W. and Vecera, S.P. (2004). Effects of Controlled Auditory-Verbal Distraction Task on Older Driver Vehicle Control. *Transportation Research Record* No. 1865. Transportation Research Board, Washington , DC.

Rizzo, M. and Dingus, T. (1996). Driving in Neurological Disease. *Neurologist*; 2: 140-160.

Rizzo, M., Anderson, S., Dawson, J., Myers, R. and Ball, K. (2000). Visual Attention Impairments in Alzheimer's Disease. *Neurology*; 54: 1954-1959.

Rizzo, M., Shi, Q., Dawson, J.D., Anderson, S.W., Kellison, I. and Pietras, T. (2005). Stop for Cops. *Transportation Research Record* No. 1922. Transportation Research Board, Washington, D.C.

Rizzo, M., Severson, J., Cremer, J. and Price, K. (2003). An Abstract Virtual Environment Tool to Assess Decision-making in Impaired Drivers. *The Second International Driving Symposium on Human Factors in Driving Assessment, Training, and Vehicle Design*, Park City, Utah, July 21-24, 2003.

Rogers, W.A. and Fisk, A.D. (1991). Age-related differences in the maintenance and modification of automatic processes: Arithmetic Stroop interference. *Human Factors*; 33: 45-56.

Romesburg, H.C. (1984). *Cluster Analysis for Researchers*. Lulu Press.

Royall, D.R., Cordes, J.A. and Polk, M. (1998). CLOX: An Executive Clock Drawing Task. *J Neurol Neurosurg Psychiatry*; 64: 588-594.

Rubin, G.S., West, S.K., Munoz, B. et al. and the Salisbury Eye Evaluation Project Team (1997). A Comprehensive Assessment of Visual Impairment in a population of older Americans. *Invest Ophthalmol Vis Sci.*; 38: 557-568.

Ruffolo, J.S., Javorsky, D.J., Tremont, G., Westervelt, H.J. and Stern, R.A. (2001). A comparison of administration procedures for the Rey-Osterrieth Complex Figure: Flowcharts Versus Pen Switching. *Psychological Assessment*; 13(3): 299-305.

Rumar, K. (1990). The Basic Driver Error: Late Detection. *Ergonomics;* 33(10/11): 1281-1290.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1987). *Akaike Information Criterion Statistics*. Dordrecht: Reidel.

Salthouse, T.A. (1985). Speed of Behavior and its Implications for Cognition. In: Birren, J.E., Schaie, K.W. (Eds.), *Handbook of the Psychology of Aging*. Van Nostrand Reinhold, New York.

Salthouse, T.A. and Kausler, D.H. (1985). Memory Methodology in Maturity. In C.J. Brainerd & M. Pressley (Eds.), *Basic processes in Memory Development* (pp. 279-311). New York: Springer-Verlag.

Salthouse, T.A., Kausler, D. and Saults, J.S. (1988). Investigation of Student Status, Background Variables, and Feasibility of Standard Tasks in Cognitive Aging Research. *Psychology and Aging*; 3(1): 29-37.

Salthouse, T.A., Mitchell, D.R.D., Skovronek, E. and Babcock, R.L. (1989). Effects of Adult Age and Working Memory on Reasoning and Spatial Abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*; 15(3): 507-516.

Salthouse, T.A., Babcock, R.L., Skovronek, E., Mitchell, D.R.D. and Palmon, R. (1990). Age and Experience Effects in Spatial Visualization. *Developmental Psychology*; 26(1): 128-136.

Salthouse, T. (1991). *Theoretical Perspectives on Cognitive Aging*. Hillsdale, NJ, Erlbaum.

Salthouse, T.A. (1992). Influence of Processing Speed on Adult Age Differences in Working Memory. *Acta Psychologica*; 79: 155-170.

Salthouse, T. and Meinz, E. (1995). Aging, Inhibition, Working memory, and Speed. *The Journals of Gerontology: Series B: Psychological Sciences and Social sciences*; 50: 297-302.

Sanders, A.F. (1970). Some aspects of the selective process in the functional visual field. *Ergonomics*; 13 (1): 101-117.

Schiff, W. and Detwiler, M.L.(1979). Information Used in Judging Impending Collision. Perception; 8: 647-658.

Schiff, W. and Oldak, R. (1990). Accuracy of Judging time to arrival: Effects of modularity, Trajectory and Sex. Journal of Experimental Psychology: *Human Perception and Performance*; 16 (2): 303-316.

Schaie, K.W. (1996). Intellectual Development in Adulthood. In *Handbook of the Psychology of Aging.* 4th Edition. pp. 266-286. Academic Press, San Diego.

Schacter, D.L. (1987). Implicit Memory: History and Current Status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*; 13: 501-518.

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall.

Schafer, J.L. and Olsen, M.K. (1998). *Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective*. The Pennsylvania State University.

Schultheis, M.T., Hillary, F. and Chute, D. L. (2003). The Neurocognitive Driving Test: Applying Technology to the Assessment of Driving Ability Following Brain Injury. *Rehabilitation Psychology*; 48(4): 275-280.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*; 6(2): 461-464.

Scialfa, C.T., Kline, D.W. and Lyman, B.J. (1987). Age Differences in Target Identification as a function of retinal location and noise level: Examination of the Useful Field of View. *Psychology and Aging*; 2(1): 14-19.

Scialfa, C.T., Thomas, D.M. and Joffe, K.M. (1994). Age Differences in Useful Field of View: An Eye Movement Analysis. *Optometry and Vision Science*; 71(12): 736-742.

Scialfa, C., Guzy, L.T., Leibowitz, H.W., Garvey, P.M., and Tyrrell, R.A. (1991). Age Differences in Estimating Vehicle Velocity. *Psychology and Aging*; 6(1): 60-66.

Scialfa, C.T., Ho, G., Caird, J.K. and Graw, T. (1999). Traffic Sign Conspicuity: The Effects of Clutter, Luminance, and Age. *Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomics Society Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society, 108-112.

Scott, A.J. and Symons, M.J. (1971). Clustering Methods Based on Likelihood Ratio Criteria. *Biometrics*; 27: 387-398.

Sekuler, A.B., Bennett, P.J. and Mamelak, M. (2000). Effects of Aging on the Useful Field of View. *Experimental Aging Research*; 26: 103-120.

Shallice, T. (1982). Specific Impairment of Planning. In D.E. Broadbent & L. Weiskrantz (Eds.), *The Neuropsychology of Cognitive Function* (pp. 199-209). London: The Royal Society.

Shinar, D. (1978). *Driver Performance and Individual Differences in Attention and Information Processing: Volume 1.* Driver Inattention. (Tech. Report DOT HS 8-801819). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Shinar, D. (1993). Traffic Safety and Individual Differences in Drivers' Attention and Information Processing Capacity. *Alcohol, Drugs and Driving*; 9: 219-237.

Shua-Haim, J.R. and Gross, J.S. (1996). A simulated driving evaluation for Patients with Alzheimer's Disease. *American Journal of Alzheimer's Disease*; 11: 2-7.

Shulman, K., Shedlesky, R. and Silver, I.L. (1986). The challenge of Time: Clock Drawing and Functioning in the Elderly. *Int Geriatr Psychiatry*; 1: 135-140.

Shulman, K.I., Gold, D.P., Cohen, C.A. and Zucchero, C.A. (1993). Clock-Drawing and Dementia in the Community: A Longitudinal Study. *International Journal of Geriatric Psychiatry*; 8: 487-496.

Shulman, K.I. (2000). Clock-Drawing: Is it the Ideal Cognitive Screening Test? *International Journal of Geriatric Psychiatry*; 15: 548-561.

Shulman, K.I., Gold, D.P., Cohen, C.A. and Zucchero, C.A. (1993). Clock Drawing and Dementia in the Community: A longitudinal Study. *International Journal of Geriatric Psychiatry*; 8: 487-496.

Shum, D.H.K., McFarland, K.A. and Bain, J.D. (1990). Construct Validity of Eight Tests of Attention: Comparison of Normal and Closed Head Injured Samples. *The Clinical Neuropsychologist*; 4: 151-162.

Sims, R.V., Owsley, C., Allman, R.M., Ball, K. and Smoot, T.M. (1998). A Preliminary assessment of the Medical and Functional Factors Associated with Vehicle Crashes in Older Adults. *Journal of the American Geriatrics Society*; 46: 556-561.

Sims, R.V., McGwin, G., Allman, R.M., Ball k. and Owsley, C. (2000). Exploratory Study of Incident Vehicle Crashes Among Older Drivers. *Journal of Gerontology: Medical Sciences*: 55A: M22-M27.

Simons, D.J. and Levin, D.T. (1997). Change Blindness. *Trends in Cognitive Sciences*; 1: 261-267.

Skaar, N., Rizzo, M. and Stierman, L. (2003). Traffic Entry Judgments by Aging Drivers. Proceedings of Driving assessment 2003. *Second International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Park City, UT, July 21-24, 2003.

Snee, R.D. (1977). Validation of Regression Models: Methods and Examples. *Technometrics*; 19(4): 415-428.

Snellgrove, C.A. (2005). *Cognitive Screening for the Safe Driving Competence of Older People with Mild Cognitive Impairment or Early Dementia*. Australian Transport Safety Bureau.

Snyder, C.H. (2005). Dementia and Driving: Autonomy Versus Safety. *Journal of the American Academy of Nurse Practitioners*; 17(10): 393-402.

Spector, P.E. (1992). *Summated Rating Scale Construction: An Introduction*. Sage Publications.

Spreen, O. and Strauss, E. (1991). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. New York: Oxford University Press.

St.Louis, E.K., McEvoy, S., Shi, Q.C. and Rizzo, M. (2005). Useful Field of View Impairment in Partial Epilepsy. Driving Assessment 2005: *3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. 06/27/2005-06/30/2005. Rockport, Maine, USA.

Stamatiadis, N., Taylor, W.C. and McKelvey F.X. (1991). Elderly Drivers and Intersection Accidents. *Transportation Quarterly*; 45(3): 377-390.

Staplin, L. (1995). Simulator and Field Measures of Driver Age Differences in Left-Turn Gap Judgements. *Transportation Research Record*, No. 1485: 49-55.

Staplin, L. and Lyles, W. (1991) Age Differences in Motion Perception and Specific Traffic maneuver Problems. *Transportation Research Record*, 1325: 23-33.

Staplin, L., Lococo, K.H., McKnight, A.J., McKnight, A.S. and Odenheimer, G.L. (1998). *Intersection Negotiation Problems of Older Drivers, Volume II: Background Synthesis on Age and Intersection Driving Difficulties.* DTNH22-93-C-05237. National Highway Traffic Safety Administration, Washington, DC.

Staplin, L., Lococo, K.H., Stewart, J. and Lawrence, E. (1999). *Safe Mobility For Older People.* Report DOT HS 808 853. National Highway Traffic Safety Administration. U.S. Department of Transportation. At http://www.nhtsa.dot.gov/people/injury/olddrive/safe/tech-doc.htm.

Staplin, L., Lococo, K.H. and Gish, K.W. (2003). *Model Driver Screening and Evaluation Program. Final Technical Report. Volume 2: Maryland Pilot Older Driver Study*. Office of Research and Traffic Records NHTSA, Washington D.C.

Stata. (2007). *Stata Multivariate Statistics Reference Manual: Release 10*. Stata Press.

Steyerberg, E.W., Harrell, F.E., Borsboom, G.J.J.M., Eijkemans, M.J.C., Vergouwe, Y. and Habbema, J.D.F. (2001). Internal Validation of predictive models: Efficiency of some procedures for Logistic Regression Analysis. *Journal of Clinical Epidemiology*; 54: 774-781.

Streiner, D.L. and Norman, G.R. (2003). *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press.

Streiner, D.L. and Norman, G.R. (2008). *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press.

Stern, E.B., Davis, E.S., Durfee, W.K., Rosenthal, T.J. and Wachtel, J. (2004). Discriminating between Brain Injured and Non-disabled Persons: A PC-based Interactive Driving Simulator Pilot Project. *Advances in Transportation Studies*: An International Journal, 2004 Special Issue.

Stolwyk, R.J., Charlton, J.L., Triggs, T.J., Iansek, R. and Bradshaw, J.L. (2006). Neuropsychological Function and Driving Ability in People with Parkinson's Disease. *Journal of Clinical and Experimental Neuropsychology*; 28: 898-913.

Strauss, E., Sherman, E.M.S. and Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. 3rd Edition. Oxford University Press.

Storandt, M., Botwinick, J., Danziger, W.L., Berg, L. and Hughes, C.P. (1984). Psychometric differentiation of mild senile dementia of the Alzheimer type. *Archives of Neurology*; 41: 497–499.

Storey, J.E., Rowland, J.T., Basic, D. and Conforti, D.A. (2001). A comparison of Five Clock Scoring Methods using ROC (receiver operating characteristics) Curve Analysis. *International Journal of Geriatric Psychiatry*; 16: 394-399.

Stutts, J. C., Stewart J.R. and Martell, C. (1998). Cognitive Test Performance and Crash Risk in an Older Driver Population. *Accident Analysis and Prevention*; 30(3): 337-346.

Suhr, J., Grace, J. and Allen, J. (1998). Quantitative and Qualitative Performance of Stroke versus Normal Elderly on Six Clock Drawing Systems. *Archives of Clinical Neuropsychology*; 13(6): 495-502.

Summala, H. (1997). Hierarchical Models of Behavioural Adaptation and Traffic Accidents. In: T. Rothengatter and E. Carbonell Vaya (editors) *Traffic and Transport Psychology, Theory and Application*. Oxford:Elsevier Science Ltd. pp. 41-52.

Summala, H. and Koivisto, I. (1990). Les Temps de Reactions de freinage de Conducteurs Subitement Sollicites: Les Conducteurs Plus Ages Conduisent plus Lentement Pour Compenser Leurs Reactions plus Lentes. *Proceedings, Driving Behaviour in a social Context*, Paris; 680-683.

Sunderland, T., Hill, J.L., Mellow, A.M., et al. (1989). Clock Drawing in Alzheimer's Disease. A Novel Measure of Dementia Severity. *Journal of American Geriatrics Society*; 37: 725-729.

Szlyk, J.P., Myers, L., Zhang, Y., Wetzel, L. and Shapiro, R. (2002). Development and Assessment of a Neuropsychological Battery to Aid in Predicting Driving Performance. *Journal of Rehabilitation Research and Development*; 39(4): 483-496.

United Nations. (2002). *Report of the Second World Assembly on Ageing*. Madrid, Spain: United Nations, April 8-12, 2002 .

Tabachnick, B.G. and Fidell, L.S. (2007). *Using Multivariate Statistics*. 5th Edition. Pearson International.

Tallman, K.S. (1992). *Driving Performance in mild Dementia*. Doctoral dissertation, University of British Columbia, Canada. (UMI No. NN80842).

Taylor, E.M. (1959). *Psychological appraisal of Children with cerebral defects*. Cambridge, MA: Harvard University Press.

Tombaugh, T.N. and Hubley, A.M. (1991). Four Studies comparing the Rey-Osterrieth and Taylor Complex figures. *Journal of clinical and experimental Neuropsychology*; 13: 587-599.

Treat, J.R., Tumbas, N.S., McDonald, S.T., Shinar, D., Hume, R.D., Mayer, R.D., Stansifer, R.L. and Castallen, N.J.(1979). *Tri-level Study of the Causes of Traffic Accidents: Final Report, Executive Summary* (Rep.No. DOT-HS-034-3-535-79-TAC(S). Washington, DC: National Highway Traffic Safety Administration.

TRB Special Report 218 (1988 a). *Demographics and Travel Trends . In Special Report 218: Transportation in an Ageing Society*, TRB, National Research Council, Washington, D.C.

TRB Special Report 218 (1988 b). *Safety of Older Persons in Traffic . In Special Report 218: Transportation in an Ageing Society*, TRB, National Research Council, Washington, D.C.

Trobe, J.D., Waller, P.F., Cook-Flannagan, C.A., et al. (1996). Crashes and Violations among Drivers with Alzheimer's Disease. *Arch Neurol.*; 53: 411-416.

Tuokko, H., Hadjistavropoulos, T., Miller, J.A. and Beattie, B.L. (1992). The Clock Test: A Sensitive Measure to Differentiate Normal Elderly from those with Alzheimer's Disease . *Journal of the American Geriatrics Society*; 40: 579-584.

Tuokko, H., Hadjistavropoulos, T., Miller, J.A., Horton, A. and Beattie, B.L. (1995). *The Clock Test.* Toronto: Multi-Health Systems.

Tuokko, H., Hadjistavropoulos, T., Rae, S. and O'Rourke, N. (2000). A Comparison of Alternative Approaches to the Scoring of Clock Drawing. *Archives of Clinical Neuropsychology*; 15(2): 137-148.

Uc, E.Y., Dawson, J.D., Rodnitzky, R.L., Smothers, J.L. and Rizzo, M. (2003). Reduction in the Useful Field of View in Parkinson's Disease. *Annals of Neurology*, Supplement; 7: S39.

Uc, E.Y., Rizzo, M., Anderson, S., Shi, Q. and Dawson, J. (2004). Route-Following and Safety Errors by Drivers with Stroke. *Transportation Research Record* No. 1899: 90-95.

Uc, E.Y., Rizzo, M., Anderson, S. W., Shi, Q. and Dawson, J.D. (2005). Driver Identification of Landmarks and Traffic Signs After a Stroke. *Transportation Research Record* No.1922: 9-14.

UK Department for Transport (2001). *Forecasting older driver accidents and casualties. Road Safety Research Report No.23*, UK Department of Transport, London.

Ulfarsson, G.F., Kim, S. and Lentz, E.T. (2006). Factors Affecting Common Vehicle-to-Vehicle Collision Types: Road Safety Priorities in an Aging Society. *Transportation Research Record* 1980: 70-78.

U.S. Census Bureau. International database. Table 094. Midyear population, by age and sex. Available at http://www.census.gov/population/www/projections/natdet-D1A.html.

Vagverket. (2001). *Trafikmedicin*. Borlange: Vagverket, Trafikmedicinska radet.

Van Zomeren, A.H., Brouwer, W.H., Rothengatter, J.A. and Snoek, J.W. (1988). Fitness to Drive a Car After Recovery From Severe Head Injury. *Arch Phys Med Rehabil*; 69: 90-96.

Van Zomeren, A.H., Brouwer, W.H. and Minderhoud, J.M. (1987). Acquired Brain Damage and Driving: A Review. *Arch Phys Med Rehab*; 68: 697-705.

Vecera, S.P. and Luck, S.J. (2002). Attention. In *Encyclopedia of the Human Brain. Vol. 1* (V.S. Ramachandran, ed.), Academic Press, San Diego, Calif., pp.269-284.

Venables, W.M. and Ripley, B.D. (2002). *Modern Applied Statistics with S*. Springer, NewYork.

Viano, D.C., Culver, C.C., Evans, L. and Frick, M. (1990). Involvement of Older Drivers in Multivehicle Side-Impact Crashes. *Accident Analysis and Prevention*; 22(2): 177-188.

Visual Awareness Inc., (2002). *UFOV Manual version 6.0.9*. Southbridge Center, 2100 Southbridge Parkway, Suite 650. Birmingham, AL 35209.

Waber, D.P. and Holmes, J.M. (1986). Assessing Children's memory productions of the Rey- Osterrieth Complex Figure. *Journal of Clinical and Experimental Neuropsychology*; 8: 565-580.

Wainer, H. (1976). Estimating Coefficients in Linear Models: It Don't Make No Nevermind. *Psychological Bulletin*; 83(2): 213-217.

Waller, J. (1967). Cardiovascular Disease, Aging, and Traffic Accidents. *J Chronic Dis*.; 20: 615-620.

Wang, C.C., Kosinski, C.J., Schwartzberg, J.G. and Shanklin, A.V. (2003). Physician's Guide to Assessing and Counseling Older Drivers. Chicago: American Medical Association. Web site: http://www.nhtsa.dot.gov/people/injury/olddrive/olderdriversbook/pages/contents.html

Wang, M.W. and Stanley, J.C. (1970). Differential Weighting: A Review of Methods and Empirical Studies. *Review of Educational Research*; 40: 663-705.

Ward, J.H. (1963). Hierarchical Groupings to Optimize an Objective Function. *Journal of American Statistical Association*; 58: 234-244.

Warm, J.S. (1984). *Sustained Attention in Human Performance*. London, Wiley.

Watson, Y.I., Arfken, C.L. and Birge, S.J. (1993). Clock Completion: An Objective Screening Test for Dementia. *Journal of the American Geriatrics Society*; 41: 1235-1240.

Wayman, J.C. (2003). Multiple Imputation for Missing Data: What Is It And How Can I Use It? *Paper presented at the 2003 Annual Meeting of the American Educational Research Association*, Chicago, IL.

Wierwille, W.W., Kieliszewski, C.A., Hanowski, R.J., Keisler, A.S. and Olson, E.C.B. (2002). *Identification and Evaluation of Older Driver Errors: Task E Report, Investigation of Critical Incidents*. FHWA, U.S. Department of Transportation.

Weisberg, S. (2005). *Applied Linear Regression*. 3rd Edition. Wiley-Interscience.

Welford, A.T. (1980). Relationships Between Reaction Time and Fatigue, Stress, Age and Sex. In: Welford, A.T., editor. *Reaction Times*. Academic Press.

Welford, A.T. (1985). Changes of Performance with Age: An Overview. In N. Charness (Ed.) *Aging and Human Performance* (pp.333-369).Chichester, UK: John Wiley & Sons, Ltd.

Welsh, K., Butters, N., Hughes, J., Mohs, R. and Heyman, A. (1991). Detection of Abnormal Memory Decline in Mild Cases of Alzheimer's Disease Using CERAD Neuropsychological Measures. *Archives of Neurology*; 48: 278-281.

Wild, K. and Cotrell, V. (2003). Identifying Driving Impairment in Alzheimer Disease: A Comparison of Self and Observer Reports versus Driving Evaluation. *Alzheimer Disease and Associated Disorders*; 17: 27-34.

Wilkinson, S.M., Fisher, L.R. and Bronfield, P. (1989). Survey of Severely Head Injured People in the Southampton Health District. *Clinical Rehabilitation*; 3; 317-328.

Withaar, F.K., Brouwer, W.H. and Van Zomeren, A.H. (2000). Fitness to Drive in Older Drivers With Cognitive Impairment. *Journal of the International Neuropsychological Society*; 6: 480-490.

Witkin, H.A., Dyk, R.D., Faterson, H.F., Goodenough, D.R. and Karp, S.A. (1962). *Psychological Differentiation: Studies of development*. New York: Wiley.

Wolf-Klein, G.P., Silverstone, F.A., Levy, A.P., Brod, M.S. and Breuer, J. (1989). Screening for Alzheimer's Disease by Clock Drawing. *Journal of the American Geriatric Association*; 37: 730-734.

Wood, J.M., Chaparro, A., Carberry, T. and Hickson, L. (2006). How Multitasking Interacts with Visual Impairment and Age on Measures of Driving Performance. *Transportation Research Record* No. 1980: 65-69.

Yaksich, S. (1985). Interaction of Older Drivers with Pedestrians in Traffic. In: Malfetti, J.L. (Ed.), *Drivers 55 + needs and Problems of older Drivers: Survey results and recommendations, proceedings of the older driver colloquium*, February 4-7, Orlando, Florida.