Development and evaluation of a Driving Observation Schedule (DOS) to study everyday driving performance of older drivers

Z. Vlahodimitrakou[1], J.L. Charlton[1], J. Langford[1], S. Koppel[1], M. Di Stefano [2], W. Macdonald [2], B. Mazer[3], I. Gelinas[3], B. Vrkljan[4], M.M. Porter[5], G.A. Smith[5], A.W. Cull[5], S. Marshall[6]

[1] Monash University Accident Research Centre, Monash University, Melbourne Australia; [2] La Trobe University, Australia; [3]McGill University, Canada; [4]McMaster University, Canada; [5]University of Manitoba, Canada; [6]University of Ottawa, Canada

Corresponding author: judith.charlton@monash.edu

Abstract

This paper describes the development and evaluation of an on-road procedure, the Driving Observation Schedule (DOS), for monitoring individual driving behavior. DOS was developed for use in the Candrive/Ozcandrive five-year prospective study of older drivers. Key features included observations in drivers' own vehicles, in familiar environments chosen by the driver, with start/end points at their own homes. Participants were 33 drivers aged 75+ years, who drove their selected route with observations recorded during intersection negotiation, lane-changing, merging, low speed maneuvers and maneuver-free driving. Driving behaviors were scored by a specialist occupational therapy driving assessor and another trained observer. Drivers also completed a post-drive survey about the acceptability of DOS. Vehicle position, speed, distance and specific roadways travelled were recorded by an in-vehicle device installed in the participant's vehicle; this device was also used to monitor participants' driving over several months, allowing comparison of DOS trips with their everyday driving. Inter-rater reliability and DOS feasibility, acceptability and ecological validity are reported here. On average, drivers completed the DOS trip in 30.48 min (SD=7.99). Inter-rater reliability measures indicated strong agreement between the trained and the expert observers: Intra-class correlations (ICC) = 0.905, CI 95% 0.747-0.965, $p < 0.0001$; Pearson product correlation, $r(18) = .83$, $p < 0.05$. Standard error of the measurement (SEM), method error (ME) and coefficient of variation (CV) measures were consistently small (3.0, 2.9 & 3.3%, respectively). Most participants reported being 'completely at ease' (82%) with the driving task and 'highly familiar with the route' (97%). Vehicle data showed that DOS trips were similar to participants' everyday driving trips in roads used, roadway speed limits, drivers' average speed and speed limit compliance. In summary, preliminary findings suggest that DOS can be scored reliably, is of feasible duration, is acceptable to drivers and representative of everyday driving. Pending further research with a larger sample and other observers, DOS holds promise as a means of quantifying and monitoring changes in older drivers' performance in environments typical of their everyday driving.

Key words: Older drivers, on-road driving, driving behavior, driving performance.

1. Introduction

1.1. Older drivers

Older drivers pose a particular challenge with respect to management of their safe mobility. They are the fastest growing sector of the driving population and a high risk age group for serious injury crashes per distance travelled (Langford & Koppel, 2006). While fragility is a major contributor to older driver crash risk (Li, Braver & Chen, 2003), the high prevalence of age-related medical conditions and impairments also contributes to their heightened risk (Marshall, 2008). This highlights the need for accurate assessments, policies and guidelines for determining fitness to drive in the context of aging, medical conditions and associated impairments.

This is a complex policy topic; not all older drivers are unsafe, and while the crash risk associated with aging may have implications for older drivers at a population level, this may not be reflected at an individual level (Anstey, Wood, Lord & Walker, 2004). For many decades, issues regarding the older driver, while by no means ignored, have been addressed in a piecemeal way. Objective data on driving behavior is a key to advancing our understanding of older driver risk and risk management, but collecting this information can pose challenges. Self-report data used for evaluating driving exposure is prone to errors due to recall or memory bias and possible capacity limitations associated with aging or medical conditions (Charlton et al., 2010). Recent studies have demonstrated discrepancies between self-reported driving behaviours and objectively derived driving data (Blanchard, Myers & Porter, 2010; Huebner, Porter & Marshall, 2006; Marshall et al., 2007; Myers, Trang & Crizzle, 2011). An additional challenge to collection of objective measures of driving performance is that using an unfamiliar vehicle and set instructions can cause test anxiety and may fail to adequately assess drivers' navigational and problem solving skills (Lundberg & Hakamies-Blomquist, 2003, Hunt, Murphy, Carr, Duchek, Buckles, & Morris, 1997; Malloon & Wood, 2004). There is an urgent need to develop test protocols which address these challenges and accurately document drivers' everyday driving performance for both research and licensing purposes.

To this end, we sought to develop a method of unobtrusive observation and evaluation of naturalistic driving using drivers' own vehicles and familiar road environments. A compelling reason for such an approach is the increasing interest in use of a 'restricted' or 'local area' license as part of a graduated de-licensing process as drivers age (Waller, 1988). We therefore sought to collect real-world observational data, highly relevant to the *usual* driving of each participant, supplemented by data from in-vehicle recording devices.

This paper describes the resultant innovative procedure for documenting individual driving behavior, using an on-road driving task, the Driving Observation Schedule (DOS). It was developed for use in the Candrive/Ozcandrive cohort study.

1.2. The Candrive/Ozcandrive cohort study

The Candrive/Ozcandrive study is a five-year multi-centre, international research program with the core objective of promoting older drivers' safe mobility. The Candrive/Ozcandrive study involves 928 drivers aged 70+ in Canada and 302 drivers aged 75+ in Australia and New Zealand. Using a longitudinal study design, the project is tracking a cohort of older drivers for up to five years, assessing changes in their functional abilities, driving patterns and driving performance. The primary purpose is to determine and validate a screening test (Decision Rule) to identify potentially at-risk drivers (Marshall, et al., 2013). Participants' driving behavior, including trip duration and distance, is recorded through an in-car recording device (ICRD) installed in the participant's own vehicle at the beginning of the study. In addition, a range of psychometric measures of functional ability, medical conditions and attitudes related to driving are documented.

The DOS will be used to study driving behaviors of the Ozcandrive cohort and a sub-sample (up to 150) of Candrive participants.

1.3. On-road driving assessment and development of DOS

Driving rehabilitation specialists generally agree that while office-based driver assessments may provide a useful component of determining fitness-to-drive, on-road assessments are the most appropriate method for determining an individual's everyday driving ability (Di Stefano & Macdonald, 2006; McCarthy, 2005) and represent the 'gold standard' for driving competence (Odenheimer, et al., 1994).

An important component of the Candrive/Ozcandrive project was the development of an on-road driving task which could be used to characterize individual drivers' everyday driving and to monitor changes in their driving performance over time. Additionally, it was expected that such a tool could supplement the (potentially rare) primary outcome measures of crashes (and police-recorded infringements/violations of traffic safety rules and regulations) for validation of the screening test.

In developing DOS, key criteria were that it should reflect drivers' everyday driving and be feasible (in light of both time and resources) to sustain within the multi-site, longitudinal study. More specifically, DOS requirements were:

- observation of ‘natural’ driving with no intervention/instruction by the observer;
- conducted in the driver’s own vehicle (since Lundberg and Hakamies-Blomquist (2003) found significant differences in driving test outcomes for drivers undertaking an on-road test using a test vehicle versus their own car);
- conducted over routes familiar to and chosen by the older driver;
- takes approximately 25-30 minutes to complete;
- rates behaviors specifically associated with older driver safety.

The Person-Environment (P-E) Fit theory of driving competence (Willis, 2000) and Michon’s Model of Driver Behavior (Michon, 1985) were used to guide development of DOS driving maneuvers and test protocol. In the context of road safety, the P-E Fit model refers to the P-E interaction that determines the individual’s ability to perform the broad array of activities essential to the task of driving adequately. Thus, for a given level of functional ability, drivers’ competency may differ as the demands of the environment change. For example, busy traffic conditions might challenge the competency of an older driver when negotiating a turn at an uncontrolled intersection, but in light traffic, the driver may be able to execute the task with no difficulty. Complementing this idea, Michon (1985) described driving behavior using a hierarchical model in which three independent performance levels were proposed to underlie the cognitive control of driving: operational (largely automatic vehicle handling); tactical (adaptive, decision-making whilst driving); and strategic (pre-drive planning) levels of control. These concepts were influential in determining the nature of the DOS task, particularly its emphasis on driver-selected, familiar routes to afford the possibility of observing drivers’ competency in environments presenting the types and levels of demand encountered in their everyday driving.

To assist with item selection and operationalization of DOS, previous older driver research was reviewed, including older driver crash epidemiology (e.g., Catchpole, et al., 2005; Fildes, et al., 1994; Korner-Bitensky, et al., 2006; Langford & Koppel, 2006), older driver self-regulatory behavior (e.g., Charlton, et al., 2006; Baldock, et al., 2006), and published driving measures (e.g., Di Stefano & Macdonald, 2003; 2006; Dobbs, et al., 1998; Galski, et al., 1992; Hunt, et al., 1997; Justiss, 2005; Kowalski, et al., 2010; Malloon & Wood, 2004; Odenheimer, et al., 1994; Ott, et al., 2012). Based on these findings, six categories of driving behaviors were identified for inclusion in the final DOS: a) Observation of Road Environment; b) Signaling; c) Speed Regulation; d) Gap Acceptance; e) Road-Rule Compliance; and f) Vehicle/Lane Positioning. The definitions for inappropriate driving behaviours are presented in Table 1. Driving maneuvers selected for inclusion in DOS were intersection negotiation, lane-changing, merging, low speed maneuvers and maneuver-free driving (i.e., straight travel path).

1.4. Purposes of this paper

The primary aims of the present study were to evaluate a new, on-road observation protocol, the Driving Observation Schedule (DOS), designed to characterize older drivers’ everyday driving behavior in their own vehicle on roads that are part of their familiar driving environment. Specifically, this paper will:

- assess the inter-rater reliability, acceptability and feasibility for ongoing use of the DOS; and
- examine the ecological validity of the DOS (how well it represented everyday driving).

2. Method

2.1 Participants

A sample of thirty-three older drivers (20 male [61%], 13 female [39%], mean age = 80.1 years, SD = 3.4, Range: 75-88 years) were recruited for the study, representing approximately ten percent of the Ozcandrive cohort. Participants were a sub-set of drivers who volunteered for the Ozcandrive study and therefore were required to meet entry criteria for the main study. Inclusion criteria were: a) aged 75 years or older; b) held a valid driver’s license; c) drove at least 4 times per week; and d) did not have an absolute contraindication to driving, as defined by the Austroads Fitness to Drive Guidelines (2006).

Participants' performance on the Mini Mental State Exam (MMSE) (Folstein et al., 1975) ranged from 25-30 (M=28.24; SD=1.39), where the maximum score representing highest performance level is 30, suggesting a relatively high level of cognitive functioning when compared with conventional benchmarks of <23 on the MMSE (cut-point for dementia; e.g., Cullen et al, 2005) and were comparable with those of the Ozcandrive cohort (M=29.0; SD=1.1).

Participants were recruited into the main study during the two-month period in which the DOS trial was conducted (July and August, 2010). With the exception of three drivers, all were subsequently enrolled into the main study. Following initial screening and assessment, three drivers were not included in the Ozcandrive study due to incompatibilities between their vehicle and the in-vehicle recording device used in the main study. All volunteers received $50 for their participation in the Ozcandrive study but received no additional compensation for participation in the DOS task.

2.2 Materials

Driving Observation Schedule

The DOS route commenced at participants' homes and was conducted on roads familiar to and chosen by participants, travelling to up to four locations within their local area. Whilst the driving route itself was not standardized, driving behaviors (both appropriate and inappropriate) were observed and documented using standardized procedures for intersection negotiation, lane-changing, merging, maneuver-free driving and low speed maneuvering. Route complexity was recorded in terms of traffic density, speed zone and number of road lanes. Inappropriate driving behaviors were coded as follows:

- observation of road environment (no mirror use; no head checking);
- signaling (incorrect use of indicator signal or failure to use signals);
- speed regulation (too fast; too slow);
- gap acceptance (missed opportunity; unsafe gap; failure to yield);
- road rule compliance (non-compliance with relevant signage, e.g., stop sign); and
- vehicle /lane position (lane drifting, hits curb, following too close).

A data dictionary and detailed instruction manual provided operational definitions and guided the coding of observations. (Refer Table 1 for extract from instruction manual relating to inappropriate behaviors which guided observer scoring.)

The observer was also required to document the occurrence of 'critical errors', defined as errors which result in: 1) the observer terminating the DOS; 2) the vehicle being involved in a crash or near-crash; and/or 3) the observer using verbal prompts either to prevent an error escalating in severity or to correct the error.

The total DOS score (maximum 100 points) was derived as: the total number of driving behaviors completed appropriately, divided by the total number observed, less one point for each error performed during maneuver-free driving, less two points for each critical error, multiplied by 100. The computation of the total DOS score was adapted from an approach commonly employed in driving assessment research (see Di Stefano & Macdonald, 2003; Odenheimer et al., 1994). Weighting of errors is controversial due to the multitude of factors that contribute to the level of "severity" assigned to a given error and the possible range of safety implications resulting from the error (Dobbs et al, 1998, Di Stefano & Macdonald, 2006). For example, Justiss and Stav (2006) and others (e.g., Dobbs et al., 1998) used a more complex rating of errors and applied heavier penalty to critical errors. It should be noted that only one critical error incident was recorded and therefore, regardless of the weighting assigned here, its contribution to overall DOS scores in the current study was minimal.

Post-Drive Survey A post-drive survey comprising four items was developed to assess drivers' perceptions of their DOS experience following its completion. Participants were asked to rate the quality of their overall driving during the DOS, the difficulty of the driving task compared with their everyday driving, their familiarity with the selected route, and their level of comfort with being observed.

Driving Behavior Data Driving-related data were recorded using an ICRD installed in the participant's own vehicle at the beginning of the study[1]. Data recorded by the ICRD included vehicle position, distance travelled and speed of travel.

Data from DOS trips and from everyday driving for a period of approximately 4 months were recorded on a memory card within the ICRD and subsequently processed using Candrive DTS software (Persentech Inc, Winnipeg, Canada). Several files were extracted: a second-by-second file of all trips; a Google Earth mapping file containing all trips; and a summary file of trip duration and distance. For current purposes, a driving 'trip' was defined from 'vehicle engine on' to 'engine off'. Prior to analyzing trip characteristics, data files were filtered to exclude: global positioning system (GPS) noise; trips taken by other drivers of the vehicle; trips to/from the Ozcandrive office; and 'trips' that had no distance associated with them (e.g. engine idling only). Details of how the data were analyzed to provide the variables of interest are reported by Smith and colleagues (2012).

For this part of the study there were data for twenty-three participants. Three participants' vehicles were not compatible with the ICRD so the device could not be installed. Data files for 7 participants were excluded from this part of the analysis; this was due to the radio frequency device not identifying other driver trips properly (n = 6), and missing data due to other technical problems (n = 1).

2.3 Procedure

Ethics approval was obtained from the Monash University Human Research Ethics Committee (MUHREC), and all participants provided written informed consent. The DOS was implemented by a trained observer (Observer A) and a specialist Occupational Therapy driving assessor (Observer B). Both observers underwent six hours of training with another occupational therapist driving expert in both classroom and on-road environments. Training included familiarization with the DOS observation and recording procedures and general principles of driving assessment. All participants were observed by Observer A and 18 participants (55%) were also observed by Observer B.

Prior to commencement of the trip, participants were asked to nominate three to four nearby locations to which they regularly drive, and to devise a driving route commencing and ending at home and linking the nominated destinations within a 25-30 minute round trip.

Throughout the drive, the Observer(s) documented driving behavior for maneuvers specified in the DOS as they occurred during the agreed route, scoring these against the DOS criteria. After approximately 20 minutes, regardless of number of locations reached, participants were asked to return home. The observations were conducted in the participant's vehicle with the observer seated in the rear left seat to ensure that critical aspects of driving behavior were observable (see Fox, et al., 1998). (When both observers were present in the vehicle, Observer B was seated in the rear left seat and Observer A was seated in the rear middle seat.)

DOS was designed to study driving behaviors that reflect everyday driving, so drivers were encouraged to behave as they normally would, including listening to the radio. They were also permitted to have their regular passenger travel with them, if they so chose, and such passengers were allowed to assist with navigation if this was their usual practice.

2.4 Analyses

Descriptive statistics were used to provide summary data for the DOS trip including duration and distance. Information re feasibility of the DOS included whether participants were able to devise a suitable route within the designated time range, and acceptability of the DOS to participant drivers was determined using responses for the post-drive survey.

Several measures were used to assess observer reliability of DOS scores (n = 18): The intra-class correlation coefficient (ICC) with 95% confidence interval (95% CI) was used to assess systematic and random error that might affect the *relative* inter-rater reliability; the Pearson product moment correlation coefficient (r) was used to assess the degree of consistency between the observers. Three forms of method error statistics were considered in order to describe the *absolute* inter-rater reliability: the standard error of the measurement (SEM), the method error (ME) and the coefficient of variation (CV) as described by Holmback and colleagues (1999). Bland-Altman plots (individual participant difference scores plotted against the mean score for both observers) were inspected for systematic bias, outliers and presence of heteroscedasticity.

Ecological validity of the DOS was determined by comparing the DOS route and everyday routes recorded by the ICRD in each participant's vehicle. Common roadways were identified by superimposing each participant's DOS route on the Google Earth map of their everyday driving routes. Percent usage was calculated for each DOS roadway in relation to everyday driving. Paired t-tests were used to assess differences in driving distance and duration between the DOS route and everyday driving trips. The proportion of time driven in specific speed categories was also calculated and repeated measures ANOVAs were used to compare DOS and everyday driving for the percent of driving time undertaken on roadways with different posted speed limits (e.g., 40, 50, 60, 70, 80, 90, 100 km/hr). Additionally, participants' compliance with posted speed limits was computed as a percentage of the total time driven above the speed limit (≥ exceeding by ≥5 km/hr and ≥10 km/hr). A repeated measures ANOVA was used to compare speed compliance in DOS and everyday driving. For all ANOVAs performed, post hoc tests were conducted using the Holm-Sidak method. Results were deemed significant when $p < 0.05$.

3. Results

3.1 DOS descriptives

All participants were able to successfully select and execute a trip that complied with the location-chaining requirement specified (up to 4 locations). The average DOS trip duration over all participants was 31 min, 30 s (SD = 7.6 min) and the majority of drivers completed within 5 minutes of the target range of 25-30 minutes (87.9%). The average distance driven was 13.8 km (SD = 5.3). Almost half of the participants nominated a shopping centre or a local shop (49%) as a location within the route. Other locations included family/friends' house (9%), medical clinic (8%), post office (5%) and library (5%).

A summary of the frequency of observed intersection negotiation, lane changing and merging maneuvers is presented in Table 2.

3.2 DOS driving performance

The average DOS score (Maximum/Optimal = 100) was relatively high (M = 88.72; SD = 7.49; Range = 72.00 - 99.50). Driving behaviors, including both appropriate and inappropriate behaviors observed during DOS, are summarized in Table 3.

3.3 Inter-Rater Reliability

The ICC value was 0.91 (CI 95% 0.747-0.965), $p < 0.0001$ and Pearson's product moment correlation was also high, $r(18) = 0.83$, $p < 0.05$, indicating a high level of inter-rater reliability and consistency between Observers. Standard error of the measurement (SEM), method error (ME) and coefficient of variation (CV) measures were consistently small (3.0, 2.9 & 3.3%, respectively) and the Bland-Altman plot revealed low levels of systematic bias with the majority of difference values (Observer A – Observer B) close to zero.

3.4 Post-drive Survey

All participants rated their DOS driving performance as 'about the same' as their everyday driving performance (100%). The majority of participants rated the difficulty of the DOS driving task as ‘about the same’ compared to their everyday driving (82%), while 12 percent rated the DOS driving task as ‘a little less difficult’ and 6 percent rated the DOS driving task as ‘a little more difficult’. Most participants reported that they were ‘highly familiar’ with the DOS route (97%). They also reported a high level of comfort with being observed during the trip. Most participants reported that they were ‘completely at ease’ (82%) and the remaining participants reported that they were ‘at ease’ (18%).

3.5 Ecological Validity

The DOS route (measured as a single round trip from/to home with up to four linked segments) ranged from 5.5 to 26.3 km (13.8 ± 5.3 km) and took 16.2 to 45.0 minutes (31.0 ± 7.6 min) to complete. Everyday driving data were collected for a period of approximately 4 months (range: 75 to 133 days). Over this time, participants completed 385 ± 155 trips, with average distance of 6.5 ± 3.0 km, and average duration of 13.8 ± 4.1 minutes. Not surprisingly, the differences in distance and duration between the DOS and everyday driving trips was significant (p’s < 0.0001).

Given the location-chaining requirements of DOS, another approach was used to compare trip duration and distances across the two settings. Each DOS trip comprised up to four sub-trips (from one location to the next). Hence, the total DOS trip distance and duration was divided by the number of sub-trips undertaken and compared with the average everyday trip. Average DOS sub-trips were significantly shorter in both distance (3.2 ± 1.2 km) and duration (7.0 ± 2.2 min) compared to average everyday driving trips (Distance: 6.1 ± 2.4 km, $t(28) = -6.284$, $p < 0.001$; Duration: 12.8 ± 3.8 min, $t(28) = -7.838$, $p < 0.0001$).

The number of roadways which were common across the DOS route and everyday driving for each participant ranged from 4 to 24. The vast majority of time during everyday driving trips was spent on roadways with speed limits of 50 km/hr (38 ± 15 %) and 60 km/hr (37 ± 16 %). The same pattern was evident for the DOS route, with values of 42 ± 16 % and 39 ± 18 %, for 50 and 60 km/hr, respectively (see Figure 1). Differences between the DOS and everyday driving trips were found in percent time on 50 km/hr roads (DOS > everyday) and on 80 km/hr roads (everyday > DOS), $p < 0.05$. No differences were found for percent time driving on 40, 60, 70, 90 and 100 km/hr speed limit roadways.

Analysis of travel speed across the two driving contexts revealed that participants drove at 40-50 km/h for a greater proportion of time during the DOS trip compared with everyday driving trips ($p < 0.05$). There were no significant differences between the DOS trip and everyday driving trips for any other speed categories (see Figure 2).

The proportion of time for which drivers exceeded the speed limit was also examined, and the mean of this was found to be greater during everyday driving trips than during the DOS trip, both for 5 km/hr and 10 km/hr over the speed limit. Differences between driving contexts were small but significant (2.6 % versus 0.8 % for 5 km/hr in excess and 1.4 % versus 0.3 % for 10km/hr in excess for everyday driving and DOS, respectively) ($p<0.05$).

4. Discussion

This paper describes the results of a study to examine the reliability, acceptability, feasibility and ecological validity of the Driving Observation Schedule (DOS). Findings showed that a suitable DOS route could be formulated from drivers’ self-nominated locations that were common destinations in everyday driving. The majority of drivers completed the DOS trip within five minutes of the targeted range of 25-30 minutes. On average, DOS trips were around 14 kilometers long and around half of the nominated locations were shopping centers.

A high level of appropriate driving behavior was observed during intersection negotiation (94%), lane changing (90%) and merging (69%). Compared with other studies (e.g. Di Stefano & Macdonald, 2003) drivers’ error rates were very low. This difference may reflect the relatively better health of the present sample compared with that of Di Stefano and Macdonald (2003), in which drivers had been referred to the licensing authority for review because their driving competence was in question. In the current study, the most frequently observed inappropriate behaviors across all driving maneuvers were signaling errors. A small proportion of behaviors were potentially more concerning for safety, including poor gap choices at intersections and inappropriate speed choices (too fast) for all three maneuver categories.

Overall, DOS scoring was found to have high levels of both relative reliability (ICC = 0.91; r = 0.85) and absolute reliability or consistency between observers' scoring (with low method error: SEM= 2.9, CV = 3.3, ME = 3.0). The strong correlations between the Occupational Therapist's and the other observer's DOS scores suggest a high level of consistency in the way that the two observers identified appropriate and inappropriate driving behaviors. The inter-rater correlation was similar to those obtained in studies assessing the reliability of standardized on-road tests (Hunt, et al., 1997; Justiss & Stav, 2006; Odenheimer et al., 1994) and considerably stronger than that reported for an unstandardized Area Driving Performance Evaluation (Janke & Eberhard, 1998). It is likely that the structured scoring system developed for the DOS contributed to the high consistency of observations. In particular, the strong inter-rater reliability is evidence of the acceptability of the DOS being conducted by personnel with specific training in this task but who are not *professionally* trained experts in driving evaluation.

Undergoing a driving 'evaluation' in which driving performance is scrutinized can often be very stressful for older individuals. Therefore, ensuring that drivers are at ease with being observed and comfortable with demands of the route is a key requirement for DOS acceptability. From a road safety viewpoint, it is also desirable that DOS trips reflect everyday driving. While the researchers were careful to emphasize to drivers that the DOS was not a test, we were conscious that the task was somewhat contrived in its location-chaining requirements (primarily to ensure that trip duration was not excessive), and that presence of an observer may cause discomfort and/or alter performance.

Results of the post-drive survey showed that all of the participants rated their overall driving during DOS as 'about the same when compared with their normal driving'. In addition, the majority of participants rated the difficulty of the driving task as 'about the same when compared to their everyday driving' and reported high familiarity with their chosen routes and ease about being observed. It seems likely that the use of drivers' own vehicles contributed to their feelings of 'ease' with the DOS procedure, although there is no direct evidence of this. Research by Lundberg and Hakamies-Blomqvist (2003) reported higher fail rates for medically-referred drivers using a test vehicle compared with drivers using their own vehicles, which the authors attributed to driving ability being compromised by drivers' need to adapt to an unfamiliar vehicle and the associated additional cognitive load. Overall, present findings suggest that the participants believed that their performance on the DOS was representative of their everyday driving, indicating that DOS has a high level of face validity. Given the increasing international interest in local area (modified) licenses (Langford & Koppel, 2011), the DOS also offers a promising approach for construction of such license tests.

DOS trips were compared with several weeks of everyday driving using data from the recording device in participants' vehicles. Results showed that the roadways used in DOS trips were also used by these older drivers in their everyday driving. Although the DOS route was longer in both duration and time than trips during everyday driving, this was expected as the DOS route was designed by combining several typical destinations ('locations') into one trip. The longer-than-normal trip durations may have caused fatigue and greater propensity for errors. However, the relatively low error rates observed would suggest that fatigue is not a large concern. Using a revised version of DOS (see Koppel, et al., 2012), it will be possible to identify the duration and distance for each separate sub-trip of the DOS route. In particular, it is expected that those sub-trips starting at the participant's home and returning from the final location to home may be more representative of everyday trips compared with other sub-trips.

The majority of driving for both DOS trips and everyday driving was on roadways with speed limits of 50 or 60 km/hr. Compared with everyday driving, DOS trips were characterized by proportionately more driving on 50 km/h roads and less on 80km/hr roads. This may be explained by the time constraints of DOS trips, which were necessarily close to drivers' homes and therefore more likely to be along residential streets with lower speed limits rather than highways with higher speed limits.

A small but significant difference in compliance with speed limits was found between the two driving contexts, with more driving over the speed limit during everyday driving. A possible explanation could be that drivers were more careful to drive at or below the speed limit when the observer(s) were in the vehicle. However, most drivers were seldom above the speed limit during everyday driving, and the extent of identified 'speeding' was very low (on average < 3 % of driving time). This is consistent with findings that older driver crashes typically do not involve high speeds (Catchpole et al., 2005). Driving speeds were also compared between DOS trips and everyday driving. Overall, there was a trend for the DOS to be characterized by slower driving than everyday driving. Significant differences were observed in percent time driving at 40-50 km/hr, with a higher percent of the DOS trip at these lower speeds compared with everyday driving. Again, a possible explanation relates to small differences in the types of roadways used in the DOS compared to everyday driving. Because DOS trips were relatively short, approximately half an hour, they would not be expected to be representative of all the driving that an older driver might undertake over several weeks. Indeed, a limitation of the 30 minute time restriction for DOS was that there was potentially less opportunity to observe drivers within high speed driving contexts (80km/hr).

Overall, results of the current study demonstrated a relatively high level of similarity between driving conditions and performance during DOS trips and everyday driving. Therefore, the DOS appears to provide an ecologically valid way to characterize and evaluate the driving of older drivers. On this basis, preliminary results suggest that DOS may be suitable for local area license testing of older drivers, pending further analysis with larger numbers to confirm these findings.

A limitation of this study was the small sample size; this sample represented approximately 10 percent of the Ozcandrive cohort. Constraints on time and resources precluded the collection of additional DOS data. Importantly, the purpose of the study was to examine the potential value of the DOS for future use in the main Ozcandrive study, so sample size needed to be sufficiently robust to address reliability and validity issues but not so large as to compromise future use of the DOS with the larger sample. It is intended that inter-rater reliability and ecological validity of DOS will be reviewed with a larger sample using the main Ozcandrive study sample and a subset of those in the Candrive study.

Another limitation of the study was the use of relatively limited ICRD data. It would be useful to validate the behavioural observations using objective measures of the same behaviours recorded by an ICRD. The ICRD used in this study provided relatively rudimentary data (speed, position, road type etc.). In future use of the DOS, covert video monitoring of driver and roadway/traffic context will be included, which will provide richer information on driver behavior and performance and enable more comprehensive assessment of the ecological validity of DOS.

Several issues relating to the coding and computation of DOS scores are also flagged for further investigation, particularly the following three. First, we acknowledge that coding of appropriate/inappropriate performance for the driving manoeuvres relied on subjective judgments of observers. However, we provided a detailed data dictionary and instruction manual to guide the coding of observations and to improve objectivity of judgments, and the strong inter-rater reliability between the two observers suggests that their judgments were closely harmonized. Second, with respect to scoring, we have noted that there is currently no uniform approach to scoring and weighting of errors. Our assignment of a two-fold weighting to critical errors over other errors was somewhat arbitrary and arguably less stringent than weightings applied by other researchers (e.g., Justiss & Stav, 2006). Given the small number of these errors (n=1) in the current study, it was not possible to review the usefulness of the weighting scheme. Future research is proposed to explore weighting of errors and implications for safety, as discussed below. Third, a limitation of the total DOS score is that it can be interpreted only in relative terms, meaning that any given driver has demonstrated more or fewer errors over time (e.g., over the duration of the Ozcandrive/Candrive study), or more/fewer errors than another driver. It is also possible to use DOS observations to describe driver performance in terms of patterns of maneuvers made and types of errors (e.g., gap acceptance, signaling). In addition, there is an opportunity for refinement of the current scoring approach to enable determination of how much a given difference in scores matters from a safety viewpoint. Recent work by Classen and colleagues (Classen, Shechtman, Awadzi, Joo, & Lanford, 2010) will be useful for advancing the measurement of driving errors and associated safety effects. Their findings demonstrated a hierarchy of importance of errors in predicting crash-related injury with highest probability for injury associated with lane maintenance, yielding, and gap acceptance errors; moderate associations with speed regulation; and vehicle positioning and adjustment to stimuli conferring a low probability. It will be important that further analyses are conducted to explore the potential refinement of the present scoring system before application to the rest of the sample.

5. Conclusions

In summary, the study showed that a suitable DOS route can be formulated for use in the Ozcandrive/Candrive project to observe and document driver performance in a way that would suggest a good level of reliability (between observers) and which could potentially be used to monitor driving performance of an individual over time. With respect to ecological validity of the DOS protocol, while some small differences were observed between DOS and everyday driving on measures of driving speed and road types, there was a high level of commonality across driving contexts. Importantly, the DOS trip was found to be acceptable to older drivers and they perceived that their driving during that trip reflected their performance in everyday driving. Future efforts will be directed towards examining inter-rater reliability with a larger sample and other observers as well as streamlining DOS using electronic data collection methods and examining its effectiveness for local area license assessments for older drivers.

Acknowledgements

The Candrive project is funded by a Team Grant from the Canadian Institutes of Health Research (CIHR) and the Ottawa Hospital Research Institute. Ozcandrive is funded by an Australian Research Council Linkage Grant (LP 100100078) to the Monash University in partnership with La Trobe University, VicRoads, Victorian Government Department of Justice, Victoria Police, Transport Accident Commission, Road Safety Trust New Zealand and Eastern Health. The authors thank Candrive/Ozcandrive research teams and older driver participants without whose valuable contribution, this research would not be possible. The driving data analyses for this project were funded in part by the Faculty of Kinesiology and Recreation Management at the University of Manitoba.

References

Anstey, K.J., Wood, J., Lord, S., Walker, J.G., 2004. Cognitive, sensory and physical factors enabling driving safety in older adults. *Clinical Psychology Review, 25*, 45-65.

Austroads, 2006. Assessing fitness to drive for commercial and private vehicle drivers. Medical standards for licensing and clinical management guidelines. A resource for health professionals in Australia. Sydney, Australia.

Baldock M.R.J., Mathias J., McLean J., Berndt A., 2006 Self-Regulation of Driving and Older Drivers' Functional Abilities. *Clinical Gerontologist 30*:53-70.

Blanchard, R.A., Myers, A.M. & Porter, M.M., 2010. Correspondence between self-reported and objective measures of driving exposure and patterns in older drivers. *Accident Analysis and Prevention, 42*, 523-529.

Catchpole, J., Styles, T., Pyta, V., Imberger, K., 2005. *Exposure and accident risk among older drivers* (Report No. 366). Australia: ARRB Group Ltd.

Charlton J.L., Oxley J., Scully J., Koppel S., Congiu M., Muir C., Fildes B., 2006. *Self-regulatory driving practices amongst older drivers in the ACT and NSW.* Monash University Accident Research Centre, Melbourne, Australia.

Charlton. J.L., Koppel, S., Odell, M., Devlin, A., Langford, J., O'Hare, M., Kopinathan, C., Andrea, D., Smith, G., Khodr, B., Edquist, J., Muir, C., Whelan, M,. 2010. Influence of chronic illness on crash involvement of motor vehicle occupants: 2nd Edition. Monash University Accident Research Centre Report No. 300, Melbourne, Australia.

Classen, S., Shechtman, O., Awadzi, K. D., Joo, Y., & Lanford, D. N. (2010). Traffic violations versus driving errors of older adults: Informing clinical practice. *American Journal of Occupational Therapy, 64*, 233–241

Di Stefano, M., Macdonald, W., 2003. Assessment of older drivers: Relationships among on-road errors, medical conditions and test outcome. *Journal of Safety Research, 34*, 415–429.

Di Stefano, M., Macdonald, W., 2006. On-the-road evaluation of driving performance. In J. Pellerito (Ed.), *Driver rehabilitation and community mobility: principles and practice,* pp. 255 - 274. St. Louis, Missouri: Elsevier Mosby.

Dobbs, A., Hellen, R., Schopflocker, D., 1998. A comparative approach to identify unsafe older drivers. *Accident Analysis and Prevention, 30*, 363-370.

Fox, G., Bowden, S., Smith, D., 1998. On-road assessment of driving competence after brain impairment: Review of current practice and recommendations for a standardized examination. *Archives of Physical Medicine and Rehabilitation, 79*, 1288-1296.

Fildes, B. N., Corben, B.F., Kent, S.M., Oxley, J., Le, T.M., Ryan, P., 1994. *Older road user crashes.* Monash University Accident Research Centre Report No. 61, Melbourne, Australia.

Galski, T., Bruno, R.L., Ehle, H.T., 1993. Prediction of behind-the-wheel driving performance in patients with cerebral brain damage: A discriminant function analysis. *The American Journal of Occupational Therapy, 47,* 391-396.

Huebner, K.D., Porter, M.M., & Marshall, S.C., 2006. Validation of an electronic device for measuring driving exposure. *Traffic Injury Prevention, 7*, 76-80.

Kowalski, K., Tuokko, H., Tallman, K., 2010.On-road evaluation: Its use for the identification of impairment and remediation of older drivers. *Physical & Occupational Therapy in Geriatrics, 28*(1), 75-85.

Holmback, A.M., Porter, M.M., Downham, D., Lexell, J., 1999 Reliability of isokinetic ankle dorsiflexor strength measurements in healthy young men and women. *Scandinavian Journal of Rehabilitation Medicine, 31,* 229–239.

Hunt, L. A., Murphy, C. F., Carr, D., Duchek, J. M., Buckles, V., Morris, J.C., 1997. 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.

Janke, M., Eberhard, J., 1998. Assessing Medically Impaired Older Drivers in a Licensing Agency Setting. *Accident Analysis and Prevention, 30*, 347-361.

Justiss M.D., 2005. Development of a behind-the-wheel driving performance assessment for older drivers (Doctor dissertation). University of Florida, America.Justiss, M. D., & Stav, W. (2006) Development of a Behind-the-Wheel Driving Performance Assessment for Older Adults, *Topics in Geriatric Rehabilitation, 22*(2), 121–128.

Koppel, S., Charlton, J.L., Langford, J., Di Stefano, M., Macdonald, W., Mazer, B., Gelinas, I., Vrkljan, B., Marshall, S., 2012. *Findings from the Candrive/Ozcandrive Study: A preliminary analysis of the Driving Observation Schedule (DOS) data.* Paper presented at the Candrive/Ozcandrive Symposium, Australian Road Safety Research, Policy and Education Conference, 4th-6th October 2012, Wellington New Zealand.

Korner-Bitensky, N., Bitensky, J., Sofer, S., Man-Son-Hing, M., Gelinas, I., 2006. Driving evaluation practices of clinicians working in the United States and Canada. *American Journal of Occupational Therapy, 60*, 428-434.

Kowalski, K., Tuokko, H., Tallman, K., 2010. On-road evaluation: Its use for the identification of impairment and remediation of older drivers. *Physical & Occupational Therapy in Geriatrics, 28*(1), 75-85.

Langford, J., Koppel, S., 2006. The case for and against mandatory age-based assessment of older drivers, *Transportation Research Part F: Traffic Psychology and Behaviour, 9*, 353-362.

Langford, J., Koppel, S., 2011. License restrictions as an under-used strategy in managing older driver safety, *Accident Analysis and Prevention, 43*, 487-493.

Li, G., Braver, E.R., Chen, L., 2003. Fragility versus excessive crash involvement as determinants of high death rates per vehicle-mile of travel among older drivers. *Accident Analysis and Prevention, 35*(2), 227-235.

Lundberg, C., Hakamies-Blomqvist, L., 2003. Driving tests with older patients: effect of unfamiliar versus familiar vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour, 6*(3), 163-173.

McCarthy, D.P., 2005. Approaches to improving elders' safe driving abilities. *Physical & Occupational Therapy in Geriatrics, 23*, 25-42.

Malloon, K., Wood, J.M., 2004. Occupational therapy assessment of open-road driving performance: Validity of directed and self-directed navigational components. *The American Journal of Occupational Therapy, 58*, 279-286.

Marshall, S.C., 2008. The role of reduced fitness to drive due to medical impairments in explaining crashes involving older drivers. Traffic Injury Prevention, 9(4), 291-298.Marshall, S.C., Molnar, F., Man-Son-Hing, Wilson, K., Stiell, I., & Porter, M.M., 2007. Measurement of driving patterns of older adults using data logging devices with and without global positioning system capability. *Traffic Injury Prevention, 8,* 260-266.

Marshall, S., Man-Son-Hing, M., Charlton, J.L., Koppel, S., Langford, J., Tuokko, H., Porter, M., Bédard, M., Vrkljan, B., Naglie, G., Rapoport, M., Korner-Bitensky, N., Gélinas, I., Mazer, B., Myers, A., Gagnon, S., & Polgar, J. (Accepted). A five year longitudinal study of older Canadian, Australian and New Zealand drivers: Methodology for the Candrive/ Ozcandrive Study. *Accident Analysis and Prevention.*

Michon, J.A., 1985. A critical view of driver behaviour models: What do we know, what should we do? In Evans, L., & Schwing, R.C., (Eds.), *Human Behaviour and Traffic Safety*. New York: Plenum Press.

Myers, A.M., Trang, A. & Crizzle, A.M., 2011. Naturalistic study of winter driving practices by older men and women: Examination of weather, road conditions, trip purposes, and comfort. Canadian *Journal on Aging, 30,* 577-589.

Odenheimer, G.L., Beaudet, M., Jette, A.M., Albert, M.S., Grande, L. & Minaker, K. L., 1994. Performance-based driving evaluation of the elderly driver: Safety, reliability, and validity. *Journal of Gerontology, 49*, 153–159.

Ott, B.R., Papandonatos, G.D., David, J.D., Barco, P., 2012. Naturalistic validation of an on-road driving test of older drivers. *Human Factors, 54*(4), 663-674.

Smith, G., Cull, A., Mence, R., Koppel, S., Charlton, J.L., Langford, J., Porter, M., 2012. The Use of a Naturalistic Driving Route for Characterizing Older Drivers. Proceedings of Canadian Multidisciplinary Road Safety Conference, 10-13 June, Banff, Canada.

Waller, P.A. (1988). Renewal licensing of older drivers. In: *Transportation in an aging society: Improving mobility and safety for older persons, Vol 2,* 72-100. Transportation Research Board, National Research Council, Washington DC.

Willis, S.L., 2000. Driving competence: The person x environment fit. In K. W. Schaie, & M. Pietrucha (Eds). Mobility and Transportation in the Elderly (pp. 269-278). New York: Springer.

Table 1 – Definitions for inappropriate driving behaviour

Driving Behaviour	Specific Error	Explanation
Observation of Road Environment: Maintaining awareness of surroundings & road environment	No Mirror Use	Non-use of rear-/ side-view mirrors
	No looking	Failure to look ahead/left/right before proceeding through intersection
Signalling: Ability to signal intention to negotiate an intersection	Inappropriate	Failure to use signal/leaving signal on after negotiating intersection/Use of incorrect signal
Speed Regulation: Adhering to posted speed limits, & regulating speed consistent with road/traffic conditions	Too Fast	Driving over speed limit or at dangerous speed for manoeuvre
	Too Slow	Driving too slowly; (consistently; a sign of overcautiousness)
Gap acceptance: Making safe judgments about presence of other vehicles & selecting a suitably risk-free point to pull into line of traffic, or cross one or more lanes of traffic	Missed Opportunity	Being overcautious/missing opportunities when selecting gap
	Unsafe Gap	Selecting unsafe gap
	Failure to Yield	Failing to yield (give right of way)
Road-Rules Compliance: Ability to follow & appropriately respond to road signs, & not cross pavement markings	Non Compliance Light/Sign	Failing to comply with road sign/traffic light
	Crossing Pavement	Crossing a pavement marking to the extent of disturbing other road users
Vehicle/Lane Positioning: Position of vehicle whilst moving or stopped, in accordance with side lane markings or relative to vehicle ahead	Out of Lane	Drifting out of lane (with or without marked lanes)
	Hitting Curb	Hitting side curb
	Inappropriate Following Distance	Driving too close to vehicle in-front

Table 2: Summary of Observed Driving Maneuvers per trip*

	Mean ± SD	Range
Intersections (Max coded = 30)	23	10 - 36
Lane Changes (Max coded = 4)	8	0 - 21
Merges (Max coded = 4)	1	0 - 2

*Note: Based on full set of scores by Observer A (n=33). Max Coded refers to limited number of maneuvers that could be recorded on the DOS score sheet

Table 3: Appropriate and Inappropriate Driving Behavior Observed during DOS

Driving Behavior	Percentage (Number)*
Intersections (Total Intersections coded)	720
Appropriate Behavior	94% (n = 674)
Inappropriate Behavior	6% (n = 46)
Frequency of particular Inappropriate Behaviors:	
Inappropriate Signaling	29
Too Fast	9
Hitting Curb	7
Unsafe Gap	3
Missing Opportunity	1
Road Rule Non-compliance	1
Lane Changing (Total Lane Changes coded)	116
Appropriate Behavior	90% (n =104)
Inappropriate Behavior	10% (n = 12)
Frequency of particular Inappropriate Behaviors:	
Inappropriate Signaling	12
Too fast	1
Merging (Total Merges coded)	32
Appropriate Behavior	69% (n =22)
Inappropriate Behavior	31% (n = 10)
Frequency of particular Inappropriate Behaviors:	
Inappropriate Signaling	7
Too fast	1

*Note based on full set of scores by Observer A (n=33).

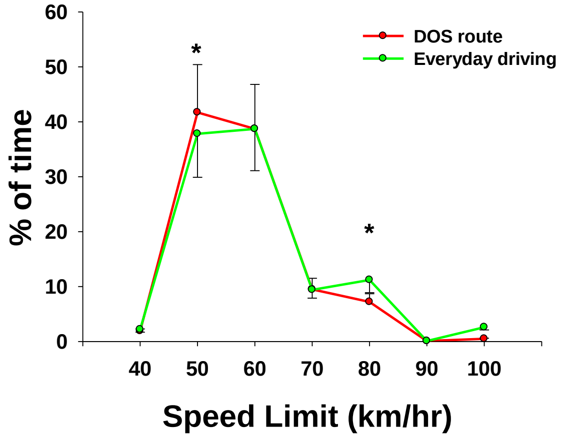

Figure 1. Percent of time driving (mean ± SE) on roadways with different speed limits for the DOS and everyday driving trips (* $p < 0.05$)

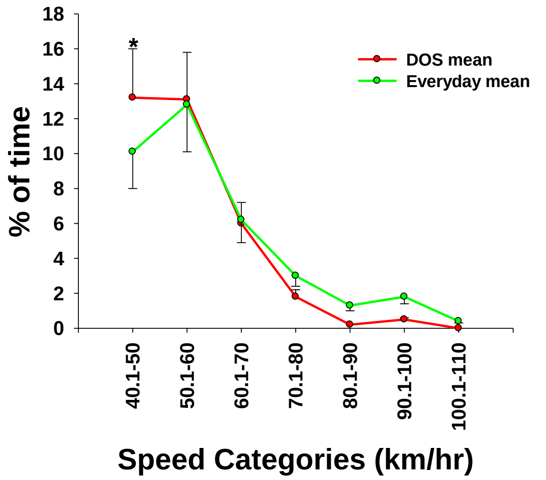

Figure 2. Percent of time (mean ± SE), spent travelling at different speeds for the DOS and everyday driving trips (* $p < 0.05$)