Stability of physical assessment of older drivers over 1 year

Andrew Smith[a], Shawn Marshall[a], Michelle Porter[b], Linda Ha[a], Michel Bédard[c], Isabelle Gélinas[d], Malcolm Man-Son-Hing[a], Barbara Mazer[d], Mark Rapoport[e], Holly Tuokko[f], Brenda Vrkljan[g]

For the Candrive research team

[a] *Ottawa Hospital Research Institute, University of Ottawa, Ottawa, Ont., Canada*

[b] *Faculty of Kinesiology and Recreation Management, University of Manitoba, Winnipeg, Man., Canada*

[c] *Centre for Research on Safe Driving, Lakehead University, Thunder Bay, Ont., Canada*

[d] *School of Physical & Occupational Therapy, McGill University, Montreal, Que., Canada*

[e] *Department of Psychiatry, University of Toronto, Toronto, Ont., Canada*

[f] *Centre on Aging, University of Victoria, Victoria, BC, Canada*

[g] *School of Rehabilitation Science, McMaster University, Hamilton, Ont., Canada*

[*] Corresponding author at: The Ottawa Hospital Rehabilitation Centre, 505 Smyth Rd., Ottawa ON Canada K1H 8M2 Tel.: (613) 737-7350 ext. 75590; fax (613) 739-6951.

E-mail address: smarshall@ottawahospital.on.ca.

Candrive research team: Shawn C. Marshall, Malcolm Man-Son-Hing, Paul Boase, Michel Bédard, Anna Byszewski, Ann B. Cranney, Hillel M. Finestone, Sylvain Gagnon, Isabelle Gélinas, Michel J. Johnson, Nicol Korner-Bitensky, Linda C. Li, Barbara L. Mazer, Frank J. Molnar, Jeannette Montufar, Anita M. Myers, Gary Naglie, Jan A. Polgar, Michelle M. Porter, Mark J. Rapoport, Ian G. Stiell, Holly A. Tuokko, Brenda H. Vrkljan, George A. Wells

This work was funded by the Canadian Institutes of Health Research (grant 90429).

The study has been registered at ClinicTrials.gov (Record #NCT01237626).

Running head: Year-long physical assessment stability

Abstract

Older adults represent the fastest-growing population of drivers with a valid driver's licence. Also common in this age group are multiple chronic medical conditions that may have an effect on physical function and driving ability. Determining the reliability of physical measures used to assess older drivers' functional ability is important to identifying those who are safe to continue driving. Most previous reliability studies of clinical physical measures of health used test–retest intervals shorter than those between patient visits with a clinician. In the present study we examined a more clinically representative interval of 1 year to determine the stability of commonly used physical measures collected during the Candrive II prospective cohort study of older drivers. Reliability statistics indicate that the sequential finger–thumb opposition, rapid pace walk and the Pelli-Robson contrast sensitivity tests have adequate stability over 1 year. Poor stability was observed for the one-legged stance and Snellen visual acuity test. Several assessments with nominal data (Marottoli method [functional neck range of motion], whispered voice test, range of motion and strength testing) lacked sufficient variability to conduct reliability analyses; however, a lack of variability between test days suggests consistency over a 1-year time frame. Our results provide evidence that specific physical measures are stable in monitoring functional ability over the course of a year.

1. Introduction

The number of adults over the age of 65 with a valid driver's licence is increasing rapidly (Bell et al., 2011; Bongaarts, 2009; Smith and Tayman, 2003). As the prevalence of chronic illnesses associated with declining functional ability as well as the driving risk in this age group are also high (Man-Son-Hing et al., 2007; Marottoli et al., 1994a; Marshall, 2008; McGwin et al., 2000; Millar, 1999), objectively determining older adults' medical fitness to drive is an important health and road safety issue. There is evidence that declines in functional ability are linked with changes in driving habits, performance, and crash rates (Margolis et al., 2002; Marottoli et al., 1994b; Marottoli et al., 2007; Sims et al., 2000; Sims et al., 2001), along with evidence that many of the impairments seen to affect older adults' driving ability are avoidable if detected and acted upon (Korner-Bitensky et al., 2011; Kua et al., 2007; Marottoli et al., 2007).

Although no single validated tool has been developed to determine medical fitness to drive, several physical assessments have been used to gain insight into drivers' functional abilities while behind the wheel of a car (Table 1). The degree to which these assessments have been examined for test–retest reliability is limited, with the majority of test–retest periods lasting no longer than 2 weeks (Table 2). While the information gathered from these studies is useful, their short durations do not reflect the longer intervals associated with patient–clinician visits (Bohannon and Schaubert, 2005) or those used in longitudinal prospective studies of older drivers' health and safety (Marshall et al., 2013 in this issue; Staplin et al., 2003). Of the studies with longer test–retest intervals (12 weeks), only those testing standard measures of strength (grip strength, knee extension) and functional ability (sit-to-stand and Timed Up & Go [TUG] tests) have involved older adults (Bohannon & Schaubert, 2005; Schaubert & Bohannon, 2005).

However, these studies are limited in that their samples are small and they did not monitor or comment on the participants' health over the testing period. Determining and monitoring the interval between testing periods becomes particularly important for longitudinal studies of older adults, as the effects of age (Alexander et al., 1991; Avlund and Davidsen, 1995; Avlund et al., 2003; Brooks and Faulkner, 1994; Skelton et al., 1994) not typically associated with analyses with short durations could affect test variables of interest. In addition, considering longitudinal investigations with study durations up to and greater than 1 year (Marshall et al., 2013 in this issue; Staplin et al., 2003), changes in lifestyle (e.g. exercise level), medication use, social status and psychological factors may influence the stability of any test variable.

Candrive II is a multicentre prospective cohort study examining the predictive validity of tools for assessing fitness to drive. Its aim is to develop an in-office screening tool that will help clinicians identify older drivers who may be unsafe to drive. A total of 928 adults aged 70 years or more have been recruited in Canada, and a further 302 (aged 75 or greater) in Australia and New Zealand (Marshall et al., 2013 in this issue). Prospective longitudinal studies like Candrive II using physical assessments as a measurement of change in health status over periods of time require test variables to be stable if no true change has occurred. Candrive II's large cohort and use of a comprehensive annual assessment of participants' sensory, physical, and cognitive function present a unique opportunity to analyse the long-term stability of commonly used physical assessments in a large group of community-dwelling older adults.

The purpose of the present study was to determine the long-term reliability of the physical measures used in Candrive II over 1 year. If the test variables are stable, older adults who remain healthy between assessments should have similar results regardless of the test–retest interval. Physical measures with moderate (0.50–0.75) or good (>0.75–1.00) correlation coefficients (Portney and Watkins, 2004) with adequate coefficients of variability (CVs) will be deemed to have moderate or good stability, which would support their use as a measure of change in physical and functional ability.

2. Methods

Long-term stability was assessed from Candrive II's baseline (year 1) and year 2 annual assessments of the Canadian participants. Assessments were performed during individual appointments lasting 2.5 to 4 hours and involved a number of easily administered measures of sensory, physical and cognitive function as well as obtaining information about health status, driving behaviours and psychosocial status. Participants were recruited from seven cities at research sites across British Columbia, Manitoba, Ontario and Quebec. Ethics approval was received by each respective research site, along with written informed consent from all participants.

2.1 Participants

To be included, participants' annual assessments must have been conducted by the same tester; 693 of the 928 participants (74.7%) met this criterion. In addition, participants had to be considered to be in stable health at both the year 1 and year 2 assessments. Only Canadian participants were used for this study. Exclusion criteria were created using data from the Expanded Cumulative Illness Rating Scale, Older American Resources and Services survey, Mini-Mental State Examination (MMSE) (Davey and Jamieson, 2004) and TUG test (Podsiadlo and Richardson, 1991), collected during the assessments. Using data from the Expanded Cumulative Illness Rating Scale, participants were excluded if they had experienced moderate symptoms requiring first-line therapy for any of the following: localized weakness, pain, problems affecting vision, hearing, cardiac or musculoskeletal health, glaucoma, macular degeneration, knee or hip replacement, stroke, seizures, Parkinson's disease, neuropathies, diabetes type I and II, or cancer. Participants were also excluded if they indicated difficulty performing activities of daily living listed in the Older American Resources and Services survey, scored less than 25 on the MMSE or had a poorer score on their year 2 assessment than on their baseline assessment on the TUG test. Because the TUG test was used as a measure to determine eligibility for the study, its stability was not tested.

2.2 Physical measures

All participants were asked to perform each of the physical measures listed in Table 1 during both their year 1 and year 2 assessments. Demographic factors such as age and gender were also recorded. All physical measures were administered by the same research associate (RA). RAs had varying levels of experience (university graduates, occupational therapists, nurses). Training of RAs included an instructional DVD explaining the assessments along with a protocol manual for reference.

2.3 Analysis

The data were analysed using SPSS version 20.0 (IBM Corporation, Armonk, NY). The relative and absolute reliability of physical measures were analysed using: a mixed-model intraclass correlation (ICC) for continuous data, the weighted kappa statistic for nominal or ordinal data, standard error of the measurement, and CV. Physical measures with moderate (0.50-0.75) or good (>0.75-1.00) correlation coefficients (Portney & Walkins 2004) have been suggested to have moderate or good stability.. A repeated-measures analysis of variance with a Bonferroni correction equal to the number of testing sessions ($p = 0.025$) was used to determine whether there was a significant difference between testing sessions.

3. Results

Of the 693 eligible participants, 91 (13.1%) were in stable health at both the year 1 and year 2 assessments. The majority of the 91 were male (66.7%); the age ranged from 70 to 89 years (M =74.6, SD = 4.52). All but two were right handed.

Results and descriptive statistics for reliability analyses of the physical measures with continuous (ICC) and ordinal (Kappa) data are presented in Table 3. All but two physical measures were found to have moderate relative repeatability (ICC 0.57–0.72). A moderate ICC value was found for both the left and right one-legged stance (ICC 0.629 and 0.590 respectively). However, a significant difference between year 1 and year 2 was observed for the left one-legged stance ($p = 0.024$), indicating a systematic improvement of participants' ability to perform the test between year 1 and year 2 (M_{year1}=15.9(±10.6), M_{year2}=18.2(±11.0)) where at year 2 the mean balance time for each leg was similar (Right leg: M_{year1}=17.6(±10.5), M_{year2}=18.2(±11.2)). This difference was accompanied by a high CV of nearly 40% for both the right and left leg. Both the Snellen test and Pelli-Robson contrast sensitivity test presented poor relative reliability between test days (k = 0.120–0.376), with the Snellen test showing very high CV values (29–60%). CVs for Pelli-Robson contrast sensitivity were much lower (less than 8%). A kappa analysis for the Pelli-Robson contrast sensitivity test for both eyes was not conducted due to insufficient variability.

Table 4 shows the proportion of participants with assessment outcomes unchanged between year 1 and 2. Up to 100% and no less than 81% of participants maintained the same result for all assessments with nominal or ordinal scales.

4. Discussion

Age-associated reductions in health increase the difficulty of any long-term analyses of test stability, as it becomes increasingly difficult to control for changes in participant health. Considering the stringent exclusion criteria, the eligibility of 91 study participants over the age of 70 years who remained in good health across the course of 1 year is a positive outcome, unlikely to be achieved if it were not for the comprehensiveness and sample size of the Candrive study.

The ICCs obtained for the one-legged stance were similar to or greater than those obtained by Luomajoki et al. (2007). However, systematic improvements for the left leg were observed along with very high CV values (almost 40%) for both the right and left leg. This indicates that for this test, typical error is very high and is not be stable over a year. It is possible that participants were more comfortable with the test during their year 2 assessment, enabling them to perform better, particularly when they only had one attempt to perform the test. It is also possible that some participants may have practised this particular task or attempted to improve their balance in the intervening year to improve their performance on this test.

A lower ICC (0.605) and a slightly higher CV (12.2%) compared to an unpublished study (ICC 0.870, CV 9.1%, Kong, 2011) was observed for the rapid pace walk. However, the previous reliability analysis compared pairs of measures collected at a 1-week interval. Despite the rigorous exclusion criteria, it is not possible to account for every factor that may influence an assessment over the course of a year; examples of such factors include aging itself, which could result in a decline in performance, and a change in lifestyle such as adoption of more walking, which could result in an improvement in performance. However, the relatively low CV indicates the measure was relatively stable.

No studies were found that examined the test–retest reliability of the sequential finger–thumb opposition and rapid foot tap tests, which made comparison difficult. Moderate ICC values and relatively low CV values for these tests suggests that the measures were sufficiently stable over the course of a year.

Few analyses of reliability were found for the Snellen test and the Pelli-Robson contrast sensitivity test, and none used a kappa analysis (Elliott et al., 1990; Lovie-Kitchin, 1988). Our results suggest that the Snellen test is not stable over extended periods. This finding is supported by “extremely poor” repeatability of measurements made using the Snellen test, with fewer than 50% of observations giving the same result over a test–retest interval of less than 3 weeks (Gibson and Sanderson, 1980). It may also be that some participants’ vision improved (new eyeglass prescription) or that vision at this age is particularly volatile and the test variable is not stable enough to test reliability at 1-year intervals. This is consistent with guideline recommendations suggesting that visual acuity be screened at least every 1 to 2 years in people over age 65 (Pelletier et al., 2009) . The CV values for the Pelli-Robson test, however, suggest that contrast sensitivity is markedly more stable over time despite low kappa values.

Similar to Schaubert and Bohannon (2005), our results were based on reliability calculations using a single measurement from each annual assessment. While multiple measurements per session are recommended (Portney and Watkins, 2004), and the benefits are undisputable for reliability analyses, given the comprehensiveness of the annual Candrive assessments (2.5 to 4 hours) it would have challenged participants' endurance to perform the physical measurements repeatedly. Our findings, like those of Schaubert and Bohannon (2005), support the view that reliable results can be obtained without multiple testing sessions/periods; however, this may also explain some discrepancies between our results and previous analyses.

Limited variability amongst test variables for the Marottoli method, whispered voice test, range of motion and strength testing made calculating a kappa reliability coefficient inappropriate. While anecdotal, this outcome supports the assertion that results between testing dates are consistent, evident when examining the proportion of participants with consistent year 1 and year 2 results. The stability of strength measures is also supported by previous findings of good reliability observed over 12 weeks (Schaubert and Bohannon, 2005).

Practically, the results from this study are more reflective of those observed in a clinical setting, in that participants were tested only once per session. In addition, the time interval better reflects follow-up visits with a clinician and those used in prospective longitudinal studies of older adults' health-related driving safety (Marshall et al., 2013 in this issue; Staplin et al., 2003). Reliability assessments from this study demonstrate that the sequential finger–thumb opposition test and rapid pace walk have moderate relative reliability and low CV values, which supports their use as a measure of change in physical and functional ability. The one-legged stance, however, while showing moderate relative reliability, is not considered stable across the course of 1 year as significant improvements between year 1 and year 2 were observed along with a CV of nearly 40%. Although both vision tests had poor relative reliability, indicated by kappa values of less than 0.4, the Pelli-Robson contrast sensitivity test appears to be the more stable of the two, with CV values less than 8% and over 85% of individuals having the same score for both eyes in years 1 and 2. Although we were unable to demonstrate it statistically, nominal data from the Marottoli method, whispered voice test, range of motion and strength testing support the view that older adults considered to be healthy at year 1 and year 2 are scored consistently. Together these findings support the stability of the sequential finger–thumb opposition, rapid pace walk, and the Pelli-Robson contrast sensitivity tests and their use in the development of a screening tool that will help clinicians and safety professionals identify older drivers who may be unsafe to drive.

Acknowledgements: We acknowledge with thanks Candrive's key partners: the National Association of Federal Retirees, Canadian Association for the Fifty-Plus (CARP), Municipal Retirees Organization Ontario, Canadian Council of Motor Transport Administrators and Transport Canada. We thank the older driver participants, without whose valuable contribution this research would not be possible. Special thanks extended to Gloria Baker for editing and preparing the manuscript for publication.

References

Alexander, N. B., Schultz, A. B., & Warwich, D. N. (1991). Rising from a chair: Effects of age functional ability on performance biomechanics. *Journals of Gerontology, 46*(3), M91-98.

American Medical Association. (2010). *Physician's guide to assessing and counseling older drivers 2nd ed.* American Medical Association.

Armstrong, A. D., MacDermid, J. C., Chinchalkar, S., Stevens, R. S., & King, G. J. (1998). Reliability of range-of-motion measurement in the elbow and forearm. *Journal of Shoulder and Elbow Surgery / American Shoulder and Elbow Surgeons ...[Et Al.], 7*(6), 573-580.

Avlund, K., & Davidsen, M. (1995). Changes in functional ability from ages 70 to 75: A danish longitudinal study. *Journal of Aging and Health, 7*(2), 254-282.

Avlund, K., Pedersen, A. N., & Schroll, M. (2003). Functional decline from age 80 to 85: Influence of preceding changes in tiredness in daily activities. *Psychosomatic Medicine, 65*(5), 771-777.

Bell, M., Wilson, T., & Charles-Edwards, E. (2011). Australia's population future: Probabilistic forecasts incorporating expert judgement. *Geographical Research, 49*(3), 261-275.

Bischoff, H. A., Stahelin, H. B., Monsch, A. U., Iversen, M. D., Weyh, A., von Dechend, M., . . . Theiler, R. (2003). Identifying a cut-off point for normal mobility: A comparison of the timed 'up and go' test in community-dwelling and institutionalised elderly women. *Age and Ageing, 32*(3), 315-320.

Bohannon, R. W., & Schaubert, K. (2005). Long-term reliability of the timed up-and-go test among community-dwelling elders. *Journal of Physical Therapy Science, 17*(2), 93-96.

Bongaarts, J. (2009). Human population growth and the demographic transition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1532), 2985-2990.

Brooks, S. V., & Faulkner, J. A. (1994). Skeletal muscle weakness in old age: Underlying mechanisms. *Medicine and Science in Sports and Exercise, 26*(4), 432-439.

Carr, D.B., Schwartzberg, J.G., Manning, L., Sempek, J., 2010. Physician's Guide to Assessing and Counseling Older Drivers, 2nd ed. National Highway Traffic Safety Association, Washington, DC.

Currie, Z., Bhan, A., & Pepper, I. (2000). Reliability of snellen charts for testing visual acuity for driving: Prospective study and postal questionnaire. *BMJ (Clinical Research Ed.), 321*(7267), 990-992.

Davey, R.J., Jamieson, S., 2004. The validity of using the mini mental state examination in NICE dementia guidelines. J. Neurol. Neurosurg. Psychiatry 75 (2), 343–344.

Eekhof, J. A. H., De Bock, G. H., De Laat, J. A. P. M., Dap, R., Schaapveld, K., & Springer, M. P. (1996). The whispered voice: The best test for screening for hearing impairment in general practice? *British Journal of General Practice, 46*(409), 473-474.

Elliott, D. B., Sanderson, K., & Conkey, A. (1990). The reliability of the pelli-robson contrast sensitivity chart. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists), 10*(1), 21-24.

Ford-Smith, C. D., Wyman, J. F., Elswick, R. K.,Jr, & Fernandez, T. (2001). Reliability of stationary dynamometer muscle strength testing in community-dwelling older adults. *Archives of Physical Medicine and Rehabilitation, 82*(8), 1128-1132. doi: 10.1053/apmr.2001.24291

Gibson, R. A., & Sanderson, H. F. (1980). Observer variation in ophthalmology. *British Journal of Ophthalmology, 64*(6), 457-460.

Hayes, K., Walton, J. R., Szomor, Z. R., & Murrell, G. A. (2001). Reliability of five methods for assessing shoulder range of motion. *The Australian Journal of Physiotherapy, 47*(4), 289-294.

Jahn, T., Cohen, R., Hubmann, W., Mohr, F., Kohler, I., Schlenker, R., . . . Schroder, J. (2006). The brief motor scale (BMS) for the assessment of motor soft signs in schizophrenic psychoses and other psychiatric disorders. *Psychiatry Research, 142*(2-3), 177-189. doi: 10.1016/j.psychres.2002.12.002

Keay, L., Munoz, B., Turano, K. A., Hassan, S. E., Munro, C. A., Duncan, D. D., . . . West, S. K. (2009). Visual and cognitive deficits predict stopping or restricting driving: The salisbury eye evaluation driving study (SEEDS). *Investigative Ophthalmology & Visual Science, 50*(1), 107-113. doi: 10.1167/iovs.08-2367

Kent-Braun, J. A., Walker, C. H., Weiner, M. W., & Miller, R. G. (1998). Functional significance of upper and lower motor neuron impairment in amyotrophic lateral sclerosis. *Muscle & Nerve, 21*(6), 762-768.

Kong, L., 2011. Reaction time and mobility testing in a cross- Canada population of older drivers. Faculty of Medicine, B.Sc. (Med) Projects, http://mspace.lib.umanitoba.ca/handle/1993/5170

Korner-Bitensky, N., Audet, T., Man-Son-Hing, M., Benoit, D., Kaizer, F., & Gelinas, I. (2011). Test-retest reliability of the preroad DriveABLE competence screen. *Physical and Occupational Therapy in Geriatrics, 29*(3), 202-212.

Kua A, Korner-Bitensky N, DesRosiers J, Man-Son-Hing M, Marshall SC. (2007). **Older driver retraining: A systematic review of evidence of effectiveness.** . *Journal of Safety Research, 38*, 81-90.

LaStayo, P. C., & Wheeler, D. L. (1994). Reliability of passive wrist flexion and extension goniometric measurements: A multicenter study. *Physical Therapy, 74*(2), 162-74; discussion 174-6.

Lovie-Kitchin, J. E. (1988). Validity and reliability of visual acuity measurements. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists), 8*(4), 363-370.

Luomajoki, H., Kool, J., de Bruin, E. D., & Airaksinen, O. (2007). Reliability of movement control tests in the lumbar spine. *BMC Musculoskeletal Disorders, 8*, 90. doi: 10.1186/1471-2474-8-90

Man-Son-Hing, M., Marshall, S. C., Molnar, F. J., & Wilson, K. G. (2007). Systematic review of driving risk and the efficacy of compensatory strategies in persons with dementia. *Journal of the American Geriatrics Society, 55*(6), 878-884. doi: 10.1111/j.1532-5415.2007.01177.x

Margolis, K. L., Kerani, R. P., McGovern, P., Songer, T., Cauley, J. A., Ensrud, K. E., & Osteoporotic, S. O. (2002). Risk factors for motor vehicle crashes in older women. *Journals of Gerontology, 57A*(1079-5006; 3), M186-M191.

Marottoli, R. A., Allore, H., Araujo, K. L., Iannone, L. P., Acampora, D., Gottschalk, M., . . . Peduzzi, P. (2007). A randomized trial of a physical conditioning program to enhance the driving performance of older persons. *Journal of General Internal Medicine, 22*(5), 590-597. doi: 10.1007/s11606-007-0134-3

Marottoli, R. A., Cooney, L. M.,Jr., Wagner, R., Doucette, J., & Tinetti, M. E. (1994a). Predictors of automobile crashes and moving violations among elderly drivers. *Annals of Internal Medicine, 121*(0003-4819; 11), 842-846.

Marottoli, R. A., Cooney, L. M.,Jr, Wagner, R., Doucette, J., & Tinetti, M. E. (1994b). Predictors of automobile crashes and moving violations among elderly drivers. *Annals of Internal Medicine, 121*(11), 842-846.

Marottoli, R. A., Richardson, E. D., Stowe, M. H., Miller, E. G., Brass, L. M., Cooney, L. M.,Jr., & Tinetti, M. E. (1998). Development of a test battery to identify older drivers at risk for self-reported adverse driving events. *Journal of the American Geriatrics Society, 46*(0002-8614; 5), 562-568.

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. doi: 10.1080/15389580801895244.

Marshall, S., Man-Son-Hing, M., Bédard, M., Charlton, J., Gagnon, S., Gélinas, I., Koppel, S., Korner-Bitensky, N., Langford, J., Mazer, B., Myers, A., Naglie, G., Polgar, J., Porter, M., Rapoport, M., Tuokko, H., Vrkljan, B., Woolnough, A., 2013. Protocol for Candrive II/Ozcandrive, a multicentre prospective older driver cohort study. Accid. Anal. Prev. 51, xxx–xxx.

McGwin, G.,Jr, Sims, R. V., Pulley, L., & Roseman, J. M. (2000). Relations among chronic medical conditions, medications, and automobile crashes in the elderly: A population-based case-control study. *American Journal of Epidemiology, 152*(5), 424-431.

Menz, H. B., Tiedemann, A., Kwan, M. M., Latt, M. D., Sherrington, C., & Lord, S. R. (2003). Reliability of clinical tests of foot and ankle characteristics in older people. *Journal of the American Podiatric Medical Association, 93*(5), 380-387.

Millar, W. J. (1999). Older drivers--a complex public health issue. *Health Reports, 11*(0840-6529; 2), 59-71.

Mullaney, M. J., McHugh, M. P., Johnson, C. P., & Tyler, T. F. (2010). Reliability of shoulder range of motion comparing a goniometer to a digital level. *Physiotherapy Theory and Practice, 26*(5), 327-333. doi: 10.3109/09593980903094230

Ng, S. S., & Hui-Chan, C. W. (2005). The timed up & go test: Its reliability and association with lower-limb impairments and locomotor capacities in people with chronic stroke. *Archives of Physical Medicine and Rehabilitation, 86*(8), 1641-1647.

Ottenbacher, K. J., Branch, L. G., Ray, L., Gonzales, V. A., Peek, M. K., & Hinman, M. R. (2002). The reliability of upper- and lower-extremity strength testing in a community survey of older adults. *Archives of Physical Medicine and Rehabilitation, 83*(10), 1423-1427.

Podsiadlo, D., & Richardson, S. (1991). The timed "up & go": A test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society, 39*(2), 142-148.

Pelletier, A. L., Thomas, J., & Shaw, F. R. (2009). Vision loss in older persons. *American Family Physician, 79*(11), 963-970.

Portney, L., Watkins, M. (2000). Reliability. In C. Mehalik (Ed.), *Foundations of clinical research: Applications to practice* (4th ed., pp. 77-94). New Jersey: Alexander, J.

SC, M. (2008). The role of reduced fitness to drive due to medical impairments in explaining crashes involving older drivers. *Traffic Inj Prev, 9*(4), 291-298.

Schaubert, K. L., & Bohannon, R. W. (2005). Reliability and validity of three strength measures obtained from community-dwelling elderly persons. *Journal of Strength and Conditioning Research, 19*(3), 717-720.

Sims, R. V., McGwin, G.,Jr., Allman, R. M., Ball, K., & Owsley, C. (2000). Exploratory study of incident vehicle crashes among older drivers. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 55*(1079-5006; 1), M22-M27.

Sims, R. V., McGwin, G.,Jr., Pulley, L. V., & Roseman, J. M. (2001). Mobility impairments in crash-involved older drivers. *Journal of Aging and Health, 13*(0898-2643; 3), 430-438.

Sinha, J. (2011). Reliability in measuring the range of motion of the aging cervical spine.

Skelton, D. A., Greig, C. A., Davies, J. M., & Young, A. (1994). Strength, power and related functional ability of healthy people aged 65-89 years. *Age and Ageing, 23*(5), 371-377.

Smith, S., & Tayman, J. (2003). An evaluation of population projections by age. *Demography, 40*(4), 741-757.

Staplin, L., Gish, K. W., & Wagner, E. K. (2003). MaryPODS revisited: Updated crash analysis and implications for screening program implementation. *Journal of Safety Research, 34*(4), 389-397. doi: 10.1016/j.jsr.2003.09.002

Swan, I. R., & Browning, G. G. (1985). The whispered voice as a screening test for hearing impairment. *The Journal of the Royal College of General Practitioners, 35*(273), 197.

Uhlmann, R. F., Rees, T. S., Psaty, B. M., & Duckert, L. G. (1989). Validity and reliability of auditory screening tests in demented and non-demented older adults. *Journal of General Internal Medicine, 4*, 90-96.

Vereeck, L., Wuyts, F., Truijen, S., & Van de Heyning, P. (2008). Clinical assessment of balance: Normative data, and gender and age effects. *International Journal of Audiology, 47*(2), 67-75. doi: 10.1080/14992020701689688

Wang, C. Y., & Chen, L. Y. (2010). Grip strength in older adults: Test-retest reliability and cutoff for subjective weakness of using the hands in heavy tasks. *Archives of Physical Medicine and Rehabilitation, 91*(11), 1747-1751. doi: 10.1016/j.apmr.2010.07.225

Youdas, J. W., Carey, J. R., & Garrett, T. R. (1991). Reliability of measurements of cervical spine range of motion--comparison of three methods. *Physical Therapy, 71*(2), 98-104; discussion 105-6.

Table 1
Summary of physical assessments and outcome measures performed during the Candrive II annual assessments.

Assessment measure	Description
Timed Up & Go test (Podsiadlo and Richardson, 1991)	Test for basic functional mobility: From a seated position, participant stands up, walks 10 feet, and returns to a seated position. Outcome is rated between 1 and 5 (1 = < 10 s normal, 2 = 10–20 s limited outdoors , 3 = 20–30 s, 4 = > 30 s, 5 = cannot do)
One-leg stance (Vereeck et al., 2008)	Test for postural steadiness: The participant stands on one leg at a time for as long as he/she can (up to 30 seconds). Outcome is recorded in seconds using a stop watch.
Marottoli method (Marottoli et al., 1998)	Test examines neck range of motion and peripheral vision: Standardized number cards are taped to a wall. The participant stands with back to the same wall and turns head to read the card out loud using peripheral vision. Similar to performing a shoulder check while driving. Outcome is pass or fail.
Whispered voice test (Swan and Browning, 1985)	Hearing test: Tester stands 0.6 m behind the participant and whispers a combination of three numbers and letters for the participant to repeat. Outcome is pass or fail.
Rapid pace walk[a] (Carr et al., 2010)	Test of lower limb functional ability: A 10-foot path is marked on the floor with tape. The participant is asked to walk the line and back. Outcome is the time in seconds to complete the test.
Sequential finger–thumb opposition (modified from original test of Jahn et al., 2006)	Test for coordination: Participant touches his/her finger to thumb in sequence and back again four times as quickly as possible. The task is repeated until done without error. Outcome is the recorded time in seconds of the successful trial.
Rapid foot taps (Kent-Braun et al., 1998)	Test for coordination: Seated, heel resting on the floor, participant taps toes as big and fast as possible in 10 seconds. Outcome is the number of foot taps within 10 seconds.
Range of motion[a] (Carr et al., 2010)	Range of motion for shoulder flexion, elbow flexion, finger curl, ankle plantar/dorsiflexion and neck rotation. Specific tests were developed based on the joint movements proposed by Carr et al. (2010), with many of them using a universal goniometer. Outcome is normal limits or not normal limits.
Muscle strength[a] (Carr et al., 2010)	Manual test of strength: Participant resists movements during ankle plantar/dorsiflexion, hip extension/flexion, shoulder abduction, hand grip and wrist flexion/extension. Outcome is rated between 0 and 5 (0 = no contraction, 1 = visible/palpable muscle contraction, no movement, 2 = movement

	with gravity eliminated, 3 = movement against gravity only, 4 = movement against gravity with some resistance, 5 = normal strength).
Snellen test[a] (Carr et al., 2010; Currie et al., 2000)	Tests visual acuity: Participant stands 10 feet from a standardized chart with 11 lines of blocked letters, the first line being very large. Subsequent rows have sequentially more numbers and letters smaller in size. Outcome is an ordinal value specified by the smallest line the participant can read without error.
Pelli-Robson contrast sensitivity[a] (Keay et al., 2009)	Tests visual contrast sensitivity: Participant stands 40 inches from a chart with 8 rows with 2 triplets of letters of uniform size and contrast. The top line has the highest contrast level (black on white). As one reads across and moves down the chart the contrast of the triplets reduces (letters become greyer) and they become more difficult to read. Participant reads out loud the letters they see until he/she reports 2 of the 3 letters within a triplet incorrectly. Outcome is an ordinal value based on the last triplet the participant reads correctly. The outcome score increases by 0.15 as triplets become more difficult to read ranging between 0.00 and 2.25.

[a] Physical measures included in the American Medical Association's Physician's Guide to Assessing and Counseling Older Drivers (Carr et al., 2010).

Table 2. Test–retest reliability and study duration for physical assessments of interest.

Assessment measure	Investigator	No. of sessions / retesting interval	Reliability coefficient[a]
One-leg stance	Luomajoki et al., 2007	2 / 2 weeks apart	k = 0.67–0.80
Whispered voice test	Uhlmann et al., 1989	2 / Same day	ICC = 0.67
	Eekhof et al., 1996	1/ same day, multiple observers	k = 0.16–1.0
Rapid pace walk	Kong, 2011	2 / 1 week apart	ICC = 0.87
Timed Up & Go test	Bohannon and Schaubert, 2005	3 / 6 weeks apart	ICC = 0.74
	Ng and Hui-Chan, 2005	2 / Within 1 week	ICC = 0.97
	Bischoff et al., 2003	3 / Same day	r_s = 0.96
Range of motion			
Shoulder flexion	Hayes et al., 2001	3 / Within 2 days	$ICC_{visual\ estimation}$ = 0.59
			$ICC_{goniometer}$ = 0.53
	Mullaney et al., 2010	2 / Same day	$ICC_{goniometer}$ = 0.91–0.97
Elbow flexion/extension	Armstrong, 1998	2 / Same day	$ICC_{goniometer}$ = 0.58–0.87
Ankle plantarflexion	Menz et al., 2003	2 / 2 weeks apart	$ICC_{protractor}$ = 0.87
Neck rotation	Sinha, 2011	2 / 1 week apart	ICC $CROM_{goniometer}$ = 0.92–0.94
	Youdas et al., 1991	3 / Same day	$ICC_{goniometer}$ = 0.78–0.90
Muscle strength			
Ankle	Ford-Smith et al., 2001	2 / 1 week apart	$ICC_{plantar\ flexion}$ = 0.61– 0.77
			$ICC_{ankle\ dorsiflexion}$ = 0.82– 0.85
	Menz et al., 2003	2 / 2 weeks apart	$ICC_{ankle\ dorsiflexion}$ = 0.88
Hip	Ford-Smith et al., 2001	2 / 1 week apart	$ICC_{hip\ extension}$ = 0.69– 0.76
			$ICC_{hip\ flexion}$ = 0.84– 0.88

Knee Shoulder abduction Hand grip Wrist flexion	Schaubert and Bohannon, 2005 Ottenbacher et al., 2002 Wang and Chen, 2010 Schaubert and Bohannon, 2005 LaStayo and Wheeler, 1994	3 / 6 weeks apart 2 / 2 days apart 2 / 1 week apart 3 / 6 weeks apart 3 / Same day	$ICC_{knee\ extension}$ = 0.92 ICC = 0.79 ICC = 0.96 ICC = 0.941–0.981 $ICC_{wrist\ flexion\ dorsal}$ = 0.92 $ICC_{wrist\ extension\ dorsal}$ = 0.84
Snellen test	Lovie-Kitchin, 1988	8 / Same day	r = 0.94
Pelli-Robson contrast sensitivity	Elliott et al., 1990	2 / 2 weeks apart	CoR = 0.15 +/- log units

Note: ICC = intraclass correlation coefficient, k = kappa coefficient, r = Pearson product-moment correlation coefficient, r_s = Spearman's rank correlation coefficient, CoR = coefficient of repeatability.

Table 3

Descriptive data and reliability coefficients of the physical measures.

Label	Year	*N*	ANOVA $p < 0.025$	Reliability	SEM	CV (%)	Mean	SD
One-leg stance (left)	1	87	0.024[a]	ICC = 0.629[b]	6.61	38.78	15.9	10.6
	2						18.2	11.0
One-leg stance (right)	1	85	0.847	ICC = 0.590[b]	6.92	38.73	17.6	10.5
	2						18.2	11.2
Rapid pace walk	1	90	0.144	ICC = 0.605[b]	0.74	12.71	6.1	1.3
	2						5.9	1.2
Sequential finger–thumb opposition (left)	1	90	0.352	ICC = 0.722[b]	2.48	10.61	13.6	2.6
	2						13.4	2.8
Sequential finger–thumb opposition (right)	1	90	0.583	ICC = 0.679[b]	1.63	12.44	13.1	2.9
	2						13.0	2.8
Rapid foot tap (left)	1	91	0.118	ICC = 0.639[b]	3.51	14.86	25.3	6.3
	2						24.4	6.4
Rapid foot tap (right)	1	91	0.072	ICC = 0.572[b]	3.80	15.23	26.3	6.5

	2						25.3	5.8
Snellen acuity both eyes)	1 2	81	0.730	k = 0.153	4.17	28.56	-	-
Snellen acuity (left eye)	1 2	85	1.000	k = 0.124	12.20	59.89	-	-
Snellen acuity (right eye)	1 2	89	0.322	k = 0.294	10.39	39.57	-	-
Pelli-Robson contrast sensitivity (left eye)	1 2	85	0.866	k = 0.376	0.14	5.69	-	-
Pelli-Robson contrast sensitivity (right eye)	1 2	85	0.322	k = 0.239	0.17	7.72	-	-

Note: SD = standard deviation, ANOVA = analysis of variance, ICC = intraclass correlation, k = kappa, SEM = standard error of the measurement, CV = coefficient of variability.
[a] Significance difference between year 1 and year 2 ($p < 0.025$).
[b] Moderate reliability (Portney and Watkins, 2004).

Table 4

Physical measures with nominal and ordinal data and insufficient variability to perform a kappa reliability analysis. A high majority of participants presented the same outcome between year 1 and year 2 assessments for the following physical measures.

Physical measure	% of participants with results unchanged between year 1 and year 2
Pelli-Robson contrast sensitivity (both eyes)	85.6
Marottoli method (left)	91.2
Marottoli method (right)	87.9
Whispered voice test (left ear)	89.0
Whispered voice test (right ear)	92.3
Range of motion (shoulder flexion, elbow flexion, finger curl, ankle plantar/dorsiflexion and neck rotation)	81.3–100*
Strength (ankle plantar/dorsiflexion, hip extension/flexion, shoulder abduction, hand grip and wrist flexion/extension)	90.1–100*

* Range includes all joints or strength tests.