Utilizing the Druid Impairment App to Assess and Enhance Senior Adults' Driving

Technical Report

William DeJong, PhD
Tufts Medical School and Impairment Science, Inc.

Michael A. Milburn, PhD
Impairment Science, Inc.

Melissa Paciulli, MCE
University of Massachusetts-Amherst, Riccio College of Engineering

Shannon C. Roberts, PhD
University of Massachusetts-Amherst, Riccio College of Engineering

Anuj K. Pradhan, PhD
University of Massachusetts-Amherst, Riccio College of Engineering

Acknowledgements

We are grateful for the guidance provided by Dr. Margie Lachman, Director of MassAITC's Aging Pilot Core, throughout our pilot project. We also thank Dr. David Paquette, Associate Research Director, MassAITC, for his patient assistance as we negotiated the University of Massachusetts' multilayered project approval process.

Note

The pilot study procedure was approved by UMass-Amherst's Institutional Review Board (IRB).

Overview

The pilot project's main goal was to validate the DRUID® impairment assessment app for use with senior adults, ages 64 to 85 years, a result that could open potential commercial markets for Impairment Science, Inc. (ISI), which developed and markets DRUID. Our study compared study participants' performance on Realtime Technology's full-scale cab driving simulator, which is housed at the Human Performance Lab (HPL), Riccio College of Engineering, University of Massachusetts-Amherst (UMass-Amherst), and their pre-simulator baseline performance on the DRUID® impairment assessment app.

DRUID has users perform four tasks that measure cognitive and motor performance by assessing balance, decision-making accuracy, reaction time, hand-eye coordination, and time estimation under conditions of divided attention. The app collects and integrates hundreds of measurements to produce an *impairment score* that ranges from 25 to 75, with a higher score indicating greater impairment.

We pioneered a novel approach: 1) developing several mini scenarios, as opposed to one, long scenario, which in turn could be divided into segments based on the participants' instructions and road conditions; and 2) analyzing dozens of outcome measures, including simulator data and a review of recorded videos.

Preliminary analyses showed that, on a few outcome measures, participants with higher baseline impairment scores were *less steady* in their driving performance, specifically with more changes in accelerator pressure and greater variability in their rate of deceleration. Other findings indicated that participants with higher impairment scores were *driving more cautiously*, with fewer lane deviations, more speedometer checks, lower maximum deceleration rates, and greater maximum braking pressure. One explanation is that, as senior adults, those who performed poorly on the DRUID app became anxious about doing well on the driving simulator and therefore became especially motivated during their 20-minute session to demonstrate exceptional driving skills.

Introduction

The pilot project's main goal was to validate the DRUID® impairment assessment app for use with senior adults, ages 64 to 85 years, a result that could open potential commercial markets for Impairment Science, Inc. (ISI), which developed and markets DRUID. As a performance-based assessment method, DRUID can detect cognitive and motor impairment due to any cause or combination of causes, including alcohol, cannabis, prescription medications, illicit drugs, fatigue, chronic disease, acute illness, injury, and concussion.

Senior adults can vary day to day in their cognitive and motor functioning. Having a reliable, valid, and easy-to-use tool to assess their current level of functioning has the potential to make significant contributions to the health, safety, and quality of life of this growing population.

Specifically, our study compared the study participants' performance on Realtime Technology's full-scale cab driving simulator, which is housed at the Human Performance Lab (HPL), Riccio College of Engineering, University of Massachusetts-Amherst (UMass-Amherst), and their pre-simulator performance on the DRUID® impairment assessment app.

Grounded in cognitive neuroscience research, DRUID has users perform four tasks that measure cognitive and motor performance by assessing balance, decision-making accuracy, reaction time, hand-eye coordination, and time estimation under conditions of divided attention (see Table 1). The app collects and integrates hundreds of measurements to produce an *impairment score* that ranges from 25 to 75, with a higher score indicating greater impairment. There are two primary versions of the app: the original Benchmark test (3 minutes) and a new Rapid test (70 seconds), which was calibrated to generate the same scores as the longer test. Both versions operate on iOS devices (iPhone, iPad) and on Android smartphones and tablets.

Table 1. Druid Benchmark's Tasks

Task 1[1]	A series of small squares and circles flash on the screen. One shape is designated as the target shape and the other as the control shape. As quickly and as accurately as possible, users are asked to touch the screen where a target shape appears and touch a small oval at the top of the screen when a control shape appears. Halfway through the task, the target and control shapes are switched.
Task 2[1]	Users press a "Start" button and then a "Stop" button when they think 30 seconds have passed. At the same time, a series of small circles flash on the screen. As quickly and accurately as possible, users are asked to touch the screen where a circle appears.
Task 3[1]	Task 3. A small circle moves around the screen and occasionally jumps a short distance. Users try to keep their finger on the circle while also counting the number of small squares that flash on the screen.
Task 4	Users stand on their right leg for 30 seconds while holding their smartphone or tablet as still as possible in their left hand, after which they stand on their left leg for 30 seconds while holding the device in their right hand.

[1] The stimuli differ within specified parameters each time a person uses the app.

Overview of Specific Aims

The first step when using DRUID is to establish a stable "unimpaired baseline" score, defined as the average of three sequential tests – taken when users are unimpaired – that are within 3.0 points of one another. In our experience, virtually all users have been able to do this successfully after 3, 4, or 5 tests, but this needed to be verified with an older population.

Aim 1a*:* Determine the percentage of participants from each age group (ages 64-74 and 75 to 85) and overall, who can establish a stable baseline score after 3, 4, 5, or more Benchmark tests.

Subsequent to submitting our proposal, ISI determined that for purposes of our research, the DRUID baseline score should be the *lowest impairment score* that a study participant can produce when first using the app. Our analysis, not reported here, showed that this was also the better procedure for our study population of senior adults.

Aim 1b: Determine what percentage of the study participants' have initial Rapid test scores that are within 3.0 points of their Benchmark baseline scores.

Aim 1c: Assess the strength of association between the study participants' initial Rapid test scores and (a) their Benchmark baseline scores and (b) the best (lowest) of their Benchmark test scores when attempting to establish their stable baseline.

The Rapid test uses abbreviated versions of the Benchmark app's tasks. After establishing their baseline score, we had originally planned to have the participants take a Rapid test to verify that the shorter test would produce comparable impairment scores for this population.

At the time we submitted our proposal, ISI had only recently developed the Rapid test. By the time we could begin collecting data, we had already demonstrated that Benchmark and Rapid test produced comparable scores. Accordingly, our study protocol did not include the Rapid test.

Aim 2a: Determine the strength of association between the study participants' driving simulator performance (total scores, plus scores for all component measures) and their Benchmark test baseline scores (total scores, plus scores for all component measures).

Aim 2b*:* Repeat these analyses using (a) the best of their Benchmark test scores when attempting to establish their stable baseline and (b) their initial Rapid test scores.

As noted, our principal goal was to validate DRUID by assessing how well the Benchmark app predicts driving simulator errors in a population of senior adults. This report includes a set of preliminary analyses that examine whether the study participants' DRUID baseline scores (as described under Aim 1a) can predict any of several outcome measures derived from their performance on the driving simulator. In the coming months, we will be conducting several additional statistical analyses.

Aim 3a: Repeat the analyses conducted for Specific Aim 2 using total impairment scores without the balance task, leaving the three tasks that measure cognitive performance.

Balance may be of less importance in predicting driving simulator performance. Shortly after we began collecting data, ISI's research and development team began exploring additional tasks for a version of DRUID that did not include the balance task. Accordingly, we will not be conducting this analysis.

Aim 3b**:** Examine whether participants' post-simulator Benchmark test scores provide evidence of fatigue-related impairment when compared to (a) their Benchmark stable baseline scores, and (b) their lowest Benchmark scores when attempting to establish their stable baseline.

We had anticipated that some of our study participants, due to the cognitive and physical demands of operating the driving simulator, might experience acute fatigue. Our preliminary analysis showed that, subsequent to using the driving simulator, some study participants showed an even lower DRUID impairment score compared to their baseline, as defined by their lowest pre-drive score. Discovering why this is the case will require further analysis.

Aim 3c: Conduct an exploratory study with 20 participants who agree to do one or more Benchmark tests each day for 21 days following their HPL study session.

Our plan was to assess the participants' compliance with the agreement to do daily testing, the number of completed tests, and whether there is a dose-response relationship between frequency of use and improved balance. At the suggestion of our advisor, Dr. Margie Lachman, we dropped this specific aim in order to concentrate on the project's other aims.

Study Procedure

Recruitment. We recruited 40 study participants via the following methods: 1) outreach emails to the Massachusetts Council on Aging (COA) and local COAs; Amherst Neighbors, which provides services and information to the town's elderly residents; Applewood at Amherst, a retirement community; and the Five College Learning in Retirement, a peer-led lifelong learning program; 2) a tabling event at the Amherst Public Library; 3) flyers posted and sent electronically to local businesses, churches, coffeehouses, community centers, libraries, senior centers, and town buildings in Amherst, Belchertown, Chicopee, Easthampton, Hadley, Longmeadow, Northampton, South Hadley, and Springfield; 4) invited presentations at events sponsored by Amherst Neighbors, the Hadley Senior Center, and the Western Massachusetts Elder Care Conference; and 5) referrals made by recruited participants.

Enrollment. Interested persons emailed the project's research associate (RA), who then reached out by telephone or email to set up a Zoom meeting to describe the study and determine their eligibility per the following criteria:

- Be 65 to 84 years of age
- Held a US driver's license for at least 10 years

- Drive at least one day per week
- Do not need any special equipment or accommodations to drive a car
- Can drive themselves to and from the appointment for the study
- Do not generally experience motion sickness in a car, either when driving or riding as a passenger

The RA also stated that participants could not be under the influence of alcohol or cannabis at the time of their appointment at the Human Performance Laboratory (HPL).

If they qualified, the RA set up their appointment and then emailed a confirmation with instructions to read an attached informed consent form (available upon request); watch a video on how to set up a Venmo account for electronic payment of $200 if they did not want to be paid in cash; and print out or download the UMass-Amherst campus map and parking instructions.

By email and text, the RA sent appointment reminders 24 hours in advance of the study participants' scheduled time, asking them to confirm their attendance, read the informed consent form, and call after arriving at the assigned parking lot so they could be escorted to the HPL. The RA telephoned the participants if they had not confirmed their attendance by the evening before their scheduled appointment.

HPL Session, Part 1: Pre-Driver Simulator. The RA showed the study participants to the driving simulator and introduced them to the HPL research staff member who would oversee that part of the study. Next, the RA presented the informed consent form and offered to review or read the form and answer questions about the study procedure before they signed two copies, one of which they would keep for their records. The RA also confirmed that they had a current Massachusetts driver's license.

Sitting at a table, the participants completed an online pre-session questionnaire on an iPad tablet (available upon request). To encourage more realistic self-ratings, the instructions reminded the participants that this information would be kept confidential. The questionnaire asked the participants to:

- Rate their overall physical health, physical balance, and emotional health during the past seven days, as well as their ability to concentrate that day, using a five-point Likert scale (1 = Very Poor, 2 = Poor, 3 = Fair, 4 = Good, and 5 = Very Good).
- State how often they have trouble falling asleep at night, staying at sleep tonight, and staying awake during the day, using a five-point Likert scale (1 = Almost never, 2 = Sometimes, 3 = Often, 4 = Very often, and 5 = Almost every night).
- Report approximately how much sleep they got during the past 24 hours, counting naps, rate the overall quality of that sleep, and state how long it had been since they last had any sleep, including naps.

- Report whether they drank a caffeinated beverage (e.g., coffee, tea, soda, energy drink) or consumed another source of caffeine (e.g., chocolate, caffeine pill) that day. If so, they stated how long it had been since they last consumed any caffeine.
- State whether they had consumed any alcohol in the past 24 hours, and if so, how many drinks they had during that time and how long it had been since their last drink.
- Indicate whether they had consumed nicotine or cannabis in the past 24 hours, and if so, how long it been since they had done so.
- Report whether they had taken any prescribed or over-the-counter medications in the past 24 hours. If so, they stated how long it been since they had done so and rated what effect they had on how alert they felt (1 = Much less alert, 2 = Less alert, 3 = No effect, 4 = More alert, and 5 = Much more alert).
- Report whether they had used other substances for recreational purposes in the past 24 hours. If so, they stated how long it been since they had done so and rated what effect they had on how alert they felt.
- State how often, as an adult, they have felt sick or nauseated when they were riding in or using a car, bus, train, airplane, small boat, ship/ferry, amusement park ride, and playground equipment (1 = Almost never, 2 = Sometimes, 3 = Often, 4 = Very often, and 5 = Almost always).
- Rate how alert they feel right now, using a nine-point scale with five labeled scale points (1 = Very alert, 3 = Alert, 5 = Neither alert nor sleepy, 7 = Sleepy but no difficulty remaining awake, and 9 = Very sleepy).
- Rate how physically comfortable they feel right now, using a 9-point scale with five labeled scale points (1 = Very comfortable, 3= Comfortable, 5 = Neither comfortable nor uncomfortable, 7 = Uncomfortable, and 9 = Very uncomfortable).
- Indicate how many days (from 0 to 7) they drive during a typical week.
- Report whether their ability to drive is limited by a health condition or by any prescribed or over-the-counter medications they take.
- State how often they try to avoid several driving situations: 1) rush hour/heavy traffic; 2) interstate/highway driving; 3) foul weather conditions; 4) nighttime driving; 5) left-hand turns against traffic; 6) rotaries/roundabouts; 7) parallel parking, and 8) any other driving situations, which they briefly described (1 = Almost never, 2 = Sometimes, 3 = Often, 4 = Very often, and 5 = Almost always).
- Report whether they had any moving violations, citations, or traffic tickets in the past 3 years. If so, they marked which infractions they had committed: 1) failure to yield, 2) driving too slowly, 3) speeding, 4) not obeying a traffic light, 5) not obeying a stop sign, 6) not obeying another traffic sign, 7) illegal right turn, 8) improper lane change, 9) not stopping at a crosswalk, 10) reckless driving, 11) texting while driving,12) phone call

while driving, 13) driving while intoxicated (DUI/DWI), 14) tailgating, 15) illegal U-turn, 16) passing a stopped school bus, 17) driving wrong way on a one-way street, 18) improper passing, and 19) any other infractions they briefly described.

- Rate how difficult it has been for them to do the following: 1) change lanes in moderate traffic, 2) turn left across multiple oncoming lanes when there is no traffic signal, 3) stay focused on driving when there are distractions (e.g., noisy passengers, radio/CD player, eating), 4) react quickly to unexpected situations such as a car backing out of a driveway, road debris, or an animal darting in front of their car, 5) pass a large vehicle (e.g., RV, tractor-trailer truck, bus) on a road without a passing lane, 6) parallel park, 7) drive at night, 8) drive when there is fog, 9) drive in an unfamiliar area, 10) drive in a large city with high-speed traffic, multiple highway interchanges, and several signs, 11) drive when there is glare or the sun is in their face, and 12) drive in a thunderstorm with heavy rains and wind (1 = Not difficult, 2 = A little difficult, 3 = Difficult, and 4 = Very difficult).

After introducing the DRUID Benchmark app, the RA invited the participants to watch a short video, developed for this project, that presented an age peer who demonstrated the app's four tasks. The video is available at https://vimeo.com/908837640/acbe5023bf?share=copy.

The RA opened DRUID, handed over the iPad tablet, and asked them to stand near (but not touching) the table for the entire test. For the balance test, the RA suggested they reach out to the table if they started to lose their balance or were afraid of falling. Alternatively, they could stand along the wall or between a nearby desk and chair, which was the option nearly all the participants chose. This situation did not arise, but if participants had full use of only one leg, they could stand on that leg for both parts of the balance test.

The participants then did three DRUID tests in a row and then up to two more tests with the goal of having two out of three sequenced tests produce impairment scores < 3.0 points apart. As noted previously, for research projects, we set the participants' baseline score as the *lowest impairment score* they can produce when first using the app. Our analysis, not reported here, showed that this was also the better procedure for this study population.

Before escorting the participants back to the driving simulator, the RA explained that the driving session was designed to minimize the possibility of anyone getting nauseous (i.e., simulator sickness), but if they did experience dizziness, nausea, or other symptoms, they should notify the HPL researcher. If this occurred, the RA would bring them back to the table and offer them water to drink while they rested. Once they felt better, they would be paid and escorted back to their car. The contact person they identified on the consent form could come for them if they needed further assistance.

Figure 1. HPL Driving Simulator

HPL Session, Part 2: Driving Simulator. We conducted the study using the high-fidelity, fixed-base, full-cab driving simulator housed in the Human Performance Laboratory (HPL) at the University of Massachusetts-Amherst (see Figure 1). The simulator utilizes the Realtime Technologies (RTI) SimCreator software engine and consists of a fully equipped 2013 Ford Fusion cab positioned in front of five projection screens that provide a 330° field of view for a highly immersive driving experience. Two dynamic side mirrors and a rearview mirror are integrated into the cab, thereby offering participants realistic rearward visibility of the simulated environment.

The simulator incorporates a five-speaker surround sound system to simulate environmental and traffic noise, along with a two-speaker in-vehicle system for replicating in-cabin auditory cues such as alerts or navigation instructions. The SimCreator engine enables the design and control of roadway environments, traffic scenarios, and edge-case events that occur outside normal operating conditions. It also allows for the presentation of alerts, visual notifications, and navigational prompts directly through the instrument cluster and center console interfaces.

Additionally, the SimADAS module equips the simulator with the capability to model advanced driver assistance system (ADAS) features such as Adaptive Cruise Control (ACC) and Traffic Jam Assist. These systems closely replicate real-world automation functionality, maintaining vehicle speed and headway according to user-defined parameters. While the simulator supports both manual driving and programmed automation through RTI's SimDriver platform that allows for automated lateral and longitudinal vehicle control, only manual driving modes were employed in the present study.

The simulator also collects comprehensive behavioral data, including vehicle control measures (e.g., speed, lane position), real-time video recordings of participants' hand and foot movements, and verbal responses.

Eye Tracker. We collected eye and head movement data using the SmartEye Pro eye-tracking system (Smart Eye AB, Gothenburg, Sweden), a high-precision, research-grade remote eye tracker designed for laboratory and simulation environments. The system is non-invasive and does not require participants to wear headsets, glasses, or other equipment, thus allowing for naturalistic driving behavior during the experiment.

The SmartEye Pro system uses multiple infrared (IR) cameras and IR flashes to track participants' eye and head movements in three dimensions. In this study, four IR cameras and corresponding IR flashes were integrated into the driving simulator cab to create a large headbox, enabling continuous tracking even when participants moved their heads or changed posture while driving. The cameras capture gaze direction, eye position, pupil diameter, and head orientation, while the IR illumination ensures robust performance under a wide range of lighting conditions.

The system operates at high sampling frequencies (up to 120 Hz) and with sub-degree accuracy for both gaze direction and head pose, ensuring precise measurement of visual attention metrics. Because of its flexible camera configuration, SmartEye Pro can maintain reliable tracking even when participants wear prescription glasses or sunglasses, as long as these do not block IR light.

The SmartEye Pro data were synchronized in real time with the simulator's kinematic and control data streams. This integration allows eye movement measures, such as point of gaze, fixation duration, and blink events, to be directly matched to driving performance metrics, including vehicle speed, steering input, lane position, and responses to traffic events. The combined dataset therefore provides a detailed, time-aligned record of both visual attention and driving behavior throughout the experimental scenarios.

Driving Simulator: Scenarios. We employed a series of short, self-contained *mini scenarios* rather than the single extended drive typically used in HPL's driving simulator experiments. Each mini scenario presented participants with a single, well-defined driving event under controlled conditions, after which the simulator environment was reset before proceeding to the next scenario. This approach served two key purposes. First, it minimized cumulative exposure to the driving simulator, important because older drivers often experience simulator sickness with longer drives. Second, this allowed us to standardize each event while reducing the need for sharp turns or prolonged complex maneuvers that can exacerbate discomfort.

Participants completed a total of eleven mini scenarios, each designed to assess driver behaviors and perception under varying roadway and traffic conditions. The first, a two-minute introductory drive on a four-lane divided highway, served as a warm-up exercise which imposed minimal demands other than maintaining lane position and speed while traffic flowed in the opposite lane.

Figure 2. Traffic Conditions

- Roadway: a) highway, b) entrance ramp, c) exit ramp, d) rural road, e) urban street, and f) residential street
- Initial number of lanes in driver's direction
- Posted speed limit: a) initial, b) number of changes, and c) final speed limit
- Number of intersections
- Traffic lights
- Stop signs (for driver and/or cross-traffic)
- Other traffic signs on driver's side
- Number of vehicles: a) directly in front of driver and b) immediately in front
- Number of vehicles: a) directly behind driver, and b) immediately behind
- Number of vehicles: a) in oncoming traffic, b) in oncoming traffic, c) adjacent to driver, and d) parked on roadway
- Number of distractions on a) driver's side of road and b) opposite side
- Number of visual barriers
- Number of pedestrians at intersection
- Number of hidden driveways

For the subsequent mini scenarios, we sought to introduce a range of traffic conditions to test the participants' driving abilities (see Figure 2). Together, these scenarios allowed us to evaluate the study participants' visual scanning, hazard perception, and decision-making skills under controlled and repeatable conditions, while also reducing overall exposure time in the simulator to minimize discomfort for older drivers.

Below is a list and brief description of the ten experimental drives we developed and analyzed. In developing the scenarios, we sought to balance the need to a) avoid features that might induce simulator sickness and b) present enough challenges to create variable driving performances. Their duration ranged from 1 to 2.5 minutes, for a total of 18 minutes (20 minutes, including the two-minute warm-up drive).

- *Scenario 1: Merging onto* Highway. The driver is instructed to proceed down an on-ramp to enter a highway.
- *Scenario 2: Tailgating.* On a rural road, the driver approaches a vehicle traveling 5-10 mph below the speed limit, which they are expected to follow at a safe distance without tailgating. Note: Unfortunately, for technical reasons, we were unable to download data from this scenario.
- *Scenario 3: Being Overtaken*. On a divided four-lane highway, a large truck overtakes the driver's vehicle on the left and remains adjacent for several seconds. The participants are expected to maintain stable lane position and, if needed, make speed adjustments.
- *Scenario 4*: *Occluding Truck*. On a residential street, the driver is instructed to pass a delivery truck parked on the right side of the road that partially obscures the view of oncoming traffic, pedestrians, and potential hazards.
- *Scenario 5*: *Hidden Driveway*. On a residential street, a vehicle emerges unexpectedly from a driveway obscured by vegetation, prompting the driver to brake or otherwise respond safely.
- *Scenario 6: Hedge at 3-Way*. On a residential street, the driver proceeds to a three-way intersection with hedges obstructing the view on the right side, which requires them to approach cautiously, stop as indicated by traffic control devices, and visually check for hazards before making a right turn.
- *Scenario 7: Sudden Yellow*. On an urban street, a traffic signal at a four-way intersection changes to yellow just seconds before the driver reaches the stop line.
- *Scenario 8: Straight Crossing Path*. On an urban street, the driver proceeds toward an intersection where a car emerged from a street on the right side without stopping and turned left. They are expected to approach the intersection cautiously to check for other vehicles that might turn without stopping or yielding.
- *Scenario 9: Left Turn at Path (1.5 minutes)*. On an urban street, the driver approaches a signalized three-way intersection and are instructed to turn left across oncoming traffic when safe.
- *Scenario 10: Gap Acceptance (2 minutes)*. The driver is instructed to make a left turn at a signalized four-way intersection on urban street, with oncoming traffic spaced at increasingly long intervals.

Driving Simulator: Data Sources. The driving sessions generated three sources of data: the driving simulator, video recordings, and the integrated eye-tracking system.

Kinematic Data from the Simulator. The driving simulator records real-time kinematic metrics that indicate vehicle dynamics and positional changes during the driving scenarios. Key variables include vehicle speed, acceleration, deceleration, heading, lane position, steering angle, and coordinates within the simulated environment. This data reveals how participants control the vehicle under various roadway conditions. For instance, lane position data can be analyzed to assess how well drivers maintained lateral stability within the traffic lanes, while speed and acceleration profiles reveal their ability to manage longitudinal vehicle dynamics, and metrics such as steering inputs and braking forces provide insights into their motor control.

Behavioral Data from Video Recordings. Simultaneous video recordings of the participants' hand, foot, and body movements help quantify their reactions to events (e.g., responding to a traffic signal change), stopping behaviors, lane-change actions (e.g., activation of turn signals, mirror-checking patterns, and execution of lane transitions), and other behavioral responses. In addition, these recordings make it possible to identify instances of incorrect or otherwise suboptimal responses that could reflect diminished situational awareness or cognitive processing.

Eye-Tracking Data. Integrated into the driving simulator cab, the SmartEye Pro system provides high-fidelity eye-tracking data to quantify visual attention and scanning patterns. Using four infrared cameras, the system records participants' eye and head movements, including metrics such as gaze duration, gaze location, gaze frequency, and pupil diameter. This dataset can be analyzed to show how participants allocate their visual attention across areas of interest (AOIs) such as roadway configurations, traffic signals, signage, other vehicles, pedestrians, potential hazards, the dashboard, and rear and sideview mirrors.

For our study of senior adults, the integration of these data provided a robust dataset for examining how they physically operated the vehicle and how they visually attended to and cognitively processed the driving environments presented in the various mini scenarios.

Driver Simulator: Procedure. The HPL researcher provided the lab's standard instructions for driving on the simulator: 1) Drive as you normally would. 2) The car does not move, and instead what you see on the screens will change. 3) When you brake and come to a complete stop, the car will stay in the same position, and the screen projection will stop moving. 4) You will hear recorded driving instructions through the speaker in the car. Next, the HPL researcher calibrated the cameras in the car.

Before beginning the experimental drives, the participants completed the introductory, two-minute drive so they would become comfortable with the simulator environment, the virtual layout of the simulated roadway, and the vehicle controls, thereby minimizing potential confusion or disorientation and ensuring their focus on the experimental mini scenarios.

We presented the experimental drives individually, with about a 20-second break after each one while the HPL researcher cued up the next drive. After completing all nine drives, the participants were escorted from the simulator and proceeded with the remainder of their session.

HPL Session, Part 3: Post-Driving Simulator. The study participants completed a single DRUID benchmark test under the same conditions as before. Next, the RA offered a snack and a bottle of water they could have while they completed a short post-session survey on the iPad tablet (available upon request).

As before, to encourage more realistic self-ratings, the instructions reminded the participants that this information would be kept confidential. The questionnaire asked the participants to:

- Provide background information about themselves:
  - Date of birth (MM/DD/YEAR)
  - Gender that best describes them: Female, Male, or Non-binary, with two additional options: “A gender not listed here,” which they would specify, and “Prefer not to say”
  - Race/ethnicity that best describes them: African American/Black, American Indian or Alaskan Native, Asian or Pacific Islander, Caucasian/White, Hispanic (Latino/Latina), or multiple races/ethnicities, with two additional options: “Other,” which they would specify, and “Prefer not to say”
  - Highest level of education: 1 = Less than a high school diploma; 2 = Completed high school (12th grade or G.E.D.); 3 = Completed vocational training; 4 = Some college after high school graduation; 5 = Associate degree; 6 = Bachelor’s degree; 7 = Some professional school after college graduation; 8 = Master’s degree; and 9 = Doctoral degree)
  - Personal income last year: 1 = $0 – $9,999; 2 = $10,000 – $24,999; 3 = 25,000 – $49,999; 4 = $50,000 – $74,999; 5 = $75,000 – $99,999; 6 = $100,000 – $149,999; 7 = $150,000+, with the additional option “Prefer not to say”
- Rate how alert they feel right now, using the same nine-point scale from the pre-session questionnaire.
- Rate how physically comfortable they feel right now, using the same nine-point scale from the pre-session questionnaire
- Indicate whether they feel 1) nauseated and 2) light-headed right now, using a five-point Likert scale (1 = No, definitely not; 5 = Yes, definitely)
- Rate how well they thought they performed on the driving simulator, using a nine-point scale with five labeled scale points (1 = Much worse than other people my age, 3 = Worse, 5 = About the same as people my age, 7 = Better, and 9 = Much better than people my age)
- Use five-point Likert scales (1 = Not at all demanding, 5 = Very demanding) to rate how mentally demanding the driving simulator and the DRUID app to be, and to rate how physically demanding they were

The RA answered any questions the participants had about the study, paid them $200 either in cash or through their Venmo account as they preferred, thanked them for their time, and escorted them to their car. The next day, the RA emailed the participants to thank them for their time and invite them to download DRUID by visiting the Impairment Science, Inc. website.

Data Analysis

We developed ten experimental mini scenarios, nine of which provided downloadable data. To commence the data analysis, we divided each scenario into segments defined by the scenario's demands. For example, Scenario 1 ("Merge Highway") asked the participants to enter an interstate highway from an entrance ramp. We created four segments:

- Segment 1 ends 30 feet before the end of the ramp. The speed limit is 40 mph; to enter the highway safely, the driver should be driving ≥ 30 mph.
- Segment 2 ends as the driver begins to cross into the highway. The speed limit is 40 mph, but to enter the highway safely, the driver should drive ≥50 mph but ≤ 75 mph.
- Segment 3 ends when the driver is told to stop and first crosses into the breakdown lane.
- Segment 4 ends when the driver stops (or the scenario reaches its endpoint)

Scenario 10 ("Gap Acceptance") asked the drivers to approach an intersection and then turn left, which required them to negotiate several oncoming vehicles. In this case, we divided the scenario into five segments:

- Segment 1 ends when the driver's car is 100 feet from the intersection.
- Segment 2 ends when the driver crosses the white line at the intersection to turn left.
- Segment 3 ends when the driver's car has cleared the intersection.
- Segment 4 ends when the driver is told to stop.
- Segment 5 ends when driver stops (or the scenario reaches its endpoint)

Descriptions of the segments for all nine mini scenarios are available upon request.

For each segment, we listed outcome measures that would exhaustively capture the participants' driving performance. Many of these measures pertain to all nine scenarios, while others are customized to fit a segment's unique driving requirements. We assigned a unique code to each type of outcome measure to facilitate analysis across segments and scenarios.

Table 2. Outcome Measures for Scenario 1, Merge onto Highway, Segment 3

Acceleration	
Greatest rate of acceleration	AC-v1
Speed a monotonically increasing function (0 = No, 1 = Yes)	AC-v2
§ Test for linear, quadratic, v. cubic function (OPTIONAL)	AC-v3
Variability in rate of acceleration (SD)	AC-v4
§ Changes in accelerator pressure (0, count)	AC-v5
§ Minimum accelerator pressure	AC-v6
Deceleration	
Greatest rate of deceleration	DC-v1
Speed a monotonically decreasing function (0 = No, 1 = Yes)	DC-v2
§ Test for linear, quadratic, v. cubic function (OPTIONAL)	DC-v3
Variability in rate of deceleration (SD)	DC-v4
§ Changes in braking pressure (0, count)	DC-v5
§ Maximum braking pressure (0, count)	DC-v6
Driving speed	
Speed upon entering highway	DS-v1
Highest speed	DS-v2
Lowest speed	DS-v3
Variability in speed (SD)	DS-v4
Percentage of time above X mph speed limit	DS-v5
§ Number of times above X mph (0, count)	DS-v6
Percentage of time below Y mph	DS-v7
§ Number of times below Y mph (0, count)	DS-v8
Speed at end of segment	DS-v1
Lateral movement	
Variability in lateral position within lane (SD)	LM-v1
Move left outside of marked lane (0, count)	LM-v2
Move right outside of marked lane (0, count)	LM-v3
Visual scanning	
Percentage of time looking straight ahead	VS-v1
Scan left (0, count)	VS-v2
Scan right (0, count)	VS-v3
Number of times fixated (> 2.0 seconds) on left side of road	VS-v4
§ Total duration (seconds)	VS-v5
Number of times fixated (> 2.0 seconds) on right side of road	VS-v6
§ Total duration (seconds)	VS-v7
Switch lanes	
Move to left lane (0 = No, 1 = Yes). *If Yes…*	SL-v1
Use left turn signal	
§ Before starting to move to left lane (0 = No, 1 = Yes)	SL-v2
§ After starting to move to left lane (0 = No, 1 = Yes)	SL-v3
Check left sideview mirror (0, count)	SL-v4
Look out left window to check blind spot (0, count)	SL-v5
Check rearview mirror (0, count)	SL-v6

Turn off left turn signal (0 = No, 1 = Yes)	SL-v7
Move back to right lane (0 = No, 1 = Yes). *If Yes…*	SL-v8
Use right turn signal	
§ Before starting to move to right lane (0 = No, 1 = Yes)	SL-v9
§ After starting to move to right lane (0 = No, 1 = Yes)	SL-v10
Check right sideview mirror (0, count)	SL-v11
Look out right window to check blind spot (0, count)	SL-v12
Check rearview mirror (0, count)	SL-v13
Turn off right turn signal (0 = No, 1 = Yes)	SL-v14

Table 2 shows several of the outcome measures for Scenario 1's Segment 3, when the driver is told to stop and first crosses into the breakdown lane. The driving simulator automatically reported most of the outcome measures, while others had to be calculated. An example from this segment is the percentage of time the participant drove below the minimum speed limit (50 mph on a highway with a maximum speed limit of 75 mph).

For other outcome measures, we had to inspect recorded videos of the participants' drives. For example, in this segment, the instructions did not ask the participants to switch to the left traffic lane after they had entered the highway, but we had to allow for the possibility that they would do so. The outcome measures highlighted in red are those requiring visual inspection.

It should be noted that there is data missing for some of the mini scenarios. First, one study participant began to feel simulator sickness near the end of the driving simulator session and asked to discontinue. Second, in other cases, some participants never experienced the relevant event because of how they drove, and therefore, there are no data to extract and retrieve. In Scenario 3, for example, as the participant drives on a divided four-lane highway, a large truck is programmed to pull alongside to the left of their car and remain adjacent for several seconds. As it turned out, 13 participants drove fast enough to prevent the truck from catching up to them.

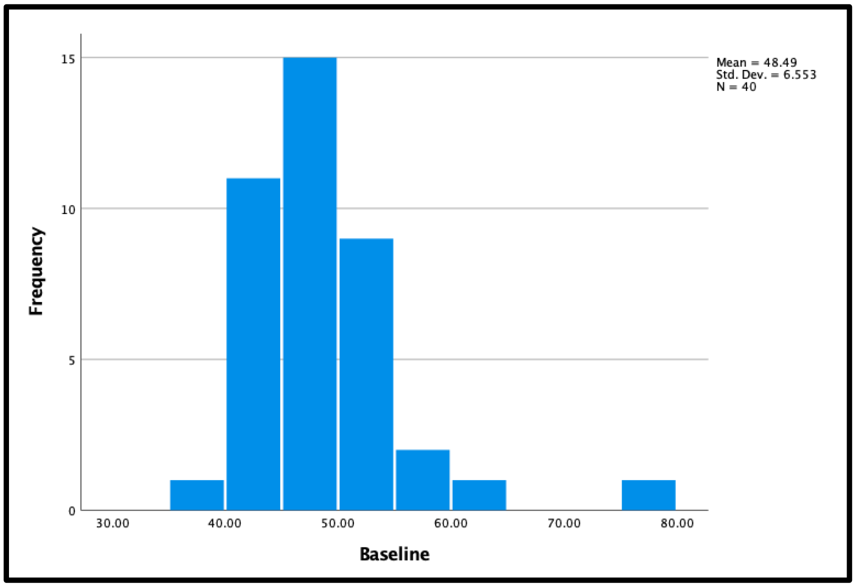
Preliminary Analyses

The participants ranged in age from 65 to 84 years, with approximately two-thirds in their 70s. Fully 92.5% were White and 57.5% were female. Ninety percent had had at least some college after high school (50.0% with a master's degree, 10.0% with a doctoral degree). Their median household income was in the $50,000-$74,999 range.

Our pilot study's principal goal was to validate DRUID for senior adults by assessing how well the Benchmark app predicts driving simulator errors. Below, we report a set of preliminary analyses that tested whether the study participants' DRUID baseline scores could predict any of several outcome measures derived from their performance on the driving simulator. In all cases, we report Pearson's correlation coefficients (r) with two-tailed tests of statistical significance.

DRUID Baseline Scores. Subsequent to submitting our proposal, Impairment Science, Inc. determined that for purposes of our research, the DRUID baseline score should be the *lowest impairment score* that a study participant can produce when first using the app. Our analysis, not reported here, showed that this was also the better procedure for our study population of senior adults. Figure 3 shows the distribution of the baseline scores.

Figure 3. Baseline DRUID Benchmark Scores (N = 40)



Before the driving simulator session, the participants used DRUID up to five times with the goal of developing a stable baseline score, defined as the average of three sequential tests that are within 3.0 points of one another. In our experience, virtually all users have been able to do this successfully after 3 to 5 tests, Accordingly, we thought that, compared to their baseline (the lowest score they had produced), the study participants might exhibit higher impairment scores due to fatigue caused by the driving simulator session.

Figure 4. Post-Session DRUID Benchmark Scores Minus Baseline Scores (N = 40)

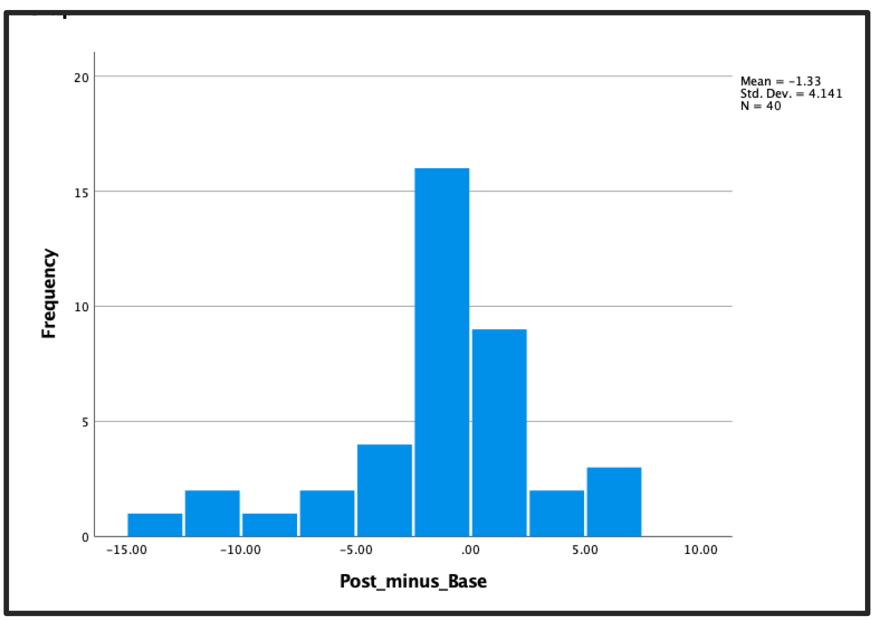


Figure 4 shows that several participants had a lower DRUID score after that session. In future analyses, we will examine possible mediators of this effect, including their baseline scores, variability across the five pre-session tests, and post-session questionnaire variables (e.g., reported alertness, how physically and mentally demanding they found DRUID and the driving simulator to be).

Two participants were unable to use DRUID effectively, with consistently high pre-session scores, well above the scores for the other participants, and produced high post-simulator scores. We dropped them from all subsequent analyses but will revisit that decision as we continue to work with the data.

Baseline Scores and Age. A consistent finding in ISI's studies is that user age is correlated with baseline DRUID scores. We have replicated that finding even though the participants' ages ranged only between 65 and 84 [r = .37, p = .029, N =35].

Baseline Scores and Hours Since Last Sleep. The more hours since the participants last slept, the lower their baseline Druid score was. This finding suggests that, for this study population, drowsiness upon waking up diminishes over time [r = -.34, p = .044, N =35].

Driving Outcomes by Scenario. With each scenario, we examined the correlations between the participants' baseline DRUID scores and the driving simulator-based outcome measures.

In all nine experimental scenarios, the baseline scores were *negatively correlated* with the percentage of time the participants looked straight ahead during the drive (VS-v1), that is, the higher their baseline impairment score, the less they looked straight ahead.

As shown below, Pearson's correlation coefficients (r) ranged between -0.32 and -0.48 and were statistically significant for six scenarios and approached significance in two others. The correlation for Scenario 3, Being Overtaken, was not significant [r = -032, p < .140].

- Scenario 1, Merging onto Highway: r = -0.35, p < .044*
- Scenario 4, Occluding Truck: r = -0.38 p < .023*
- Scenario 5, Hidden Driveway: r = -0.40, p < .023*
- Scenario 6, Hedge at 3-Way: r = -0.48, p < .004*
- Scenario 7, Sudden Yellow: r = -0.44, p < .009*
- Scenario 8, Straight Crossing Path: r = -0.33, p < .054
- Scenario 9, Left Turn at Path: r = -0.33, p < .058
- Scenario 10, Gap Acceptance: r = -037, p < .029*

We will conduct additional scenario-by-scenario analyses to determine if there were particular road conditions that created this effect.

This set of analyses revealed four additional significant findings. In Scenario 4, Occluding Truck, higher baseline DRUID scores predicted a lower maximum braking force when stopping the car at the end of the drive (ST-v9-OUT): r = -.38, p = .021, N=36. In Scenario 9, Left Turn at Path, higher baseline DRUID scores predicted a greater percentage of time above the speed limit (DS-v4): *r* = .47, *p* = .003, N = 37; and a greater number of times checking the speedometer (SA-v2): *r* = .42. *p* = .01, N =37. In Scenario 10, Gap Acceptance, higher baseline DRUID scores predicted a higher maximum rate of deceleration when stopping the car at the end of the drive (ST-v10): *r* = .42, *p* = .02, N = 29.

Table 3. Outcomes for All Scenarios Combined: Driving Simulator Data

Measure	Code	r	p	N	Higher Baseline Scores Predict
Changes in accelerator pressure (0, count)	AC-v5	.41	.011	38	More changes in accelerator pressure (AC = Acceleration)
Move left outside of marked lane (0, count)	LM-v2	-.33	.042	38	Fewer lane deviations (LM = Lateral Movement)
Check speedometer (0, count)	SA-v2	.34	.036	38	More speedometer checks (SA = Situational Awareness)
Greatest rate of deceleration	CD-v6	-0.43	.008	36	Lower greatest rate of deceleration (CD = Car in Driveway, Scenario 6)
Variability in rate of deceleration	CD-v9	.35	.039	36	Greater variability in the rate of deceleration
Maximum braking pressure	CD-v13	.41	.012	36	Greater maximum braking pressure
Changes in accelerator pressure (0, count)	SS-v9	.46	.004	38	More changes in accelerator pressure (SS = STOP SIGN)
Changes in accelerator pressure (0, count)	LT-v7	.52	<.001	38	More changes in accelerator pressure (LT = Left Turn)

Driving Outcomes for All Scenarios Combined. As noted previously, most of the driving simulator measures pertain to multiple scenarios, while others are customized to fit a segment's unique driving requirements. We assigned a unique code to each type of outcome measure to facilitate analysis across segments and scenarios.

Table 3 shows the eight statistically significant correlations between the participants' baseline DRUID scores and several driving simulator-based outcome measures. Three of the outcome measures are specific to Scenario 6, Hidden Driveway, when a vehicle emerges unexpectedly from a driveway obscured by vegetation, prompting the driver to brake or otherwise respond safely (CD-v6, CD-v9, and CD-v13).

Four of the findings indicate that participants with higher baseline impairment scores were *less steady* in their driving performance, with more changes in accelerator pressure and greater variability in the rate of deceleration.

The other findings indicate that participants with higher impairment scores were *driving more cautiously*, with fewer lane deviations, more speedometer checks, lower maximum deceleration rates, and greater maximum braking pressure. One explanation is that, as senior adults, those who performed poorly on the app became anxious about doing well on the driving simulator and therefore became especially motivated to demonstrate exceptional driving skills during their 20-minute session. Further analyses are needed to explore this possibility.

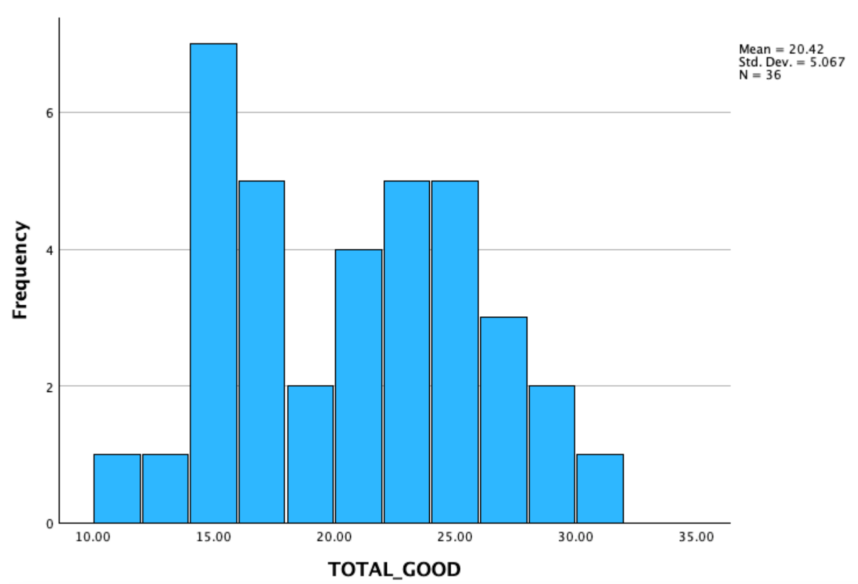
Good Driving Points. As noted previously, for other outcome measures we inspected recorded videos of the participants' driving. Watching each video multiple times, HPL researchers documented whether the participants followed a predetermined list of safe driving practices or made any critical errors.

In Scenario 7, for example, Segment 2 begins when a stop sign at a three-way intersection first becomes visible from behind a hedge. The instructions given over the speaker instruct the participants to turn right at the intersection. The segment ends when they hear the instruction to stop the car.

In responding to the stop sign, the participants should do the following: 1) decrease pressure on the accelerator; 2) apply the brake; 3) come to a complete stop before but not too far from the white stop line at the stop sign; 4) engage their right turn signal before reaching the stop sign; 5) after stopping, look both left and right for cross traffic; 6) initiate the right turn as instructed; and 7) turn into the right lane without straddling the median line. A critical error, for example, would be pressing the accelerator instead of the brake when initiating their stop.

The HPL researchers assigned a score to each item on the list (0 = No, 1 = Yes). Our preliminary analysis examined the total number of "good driving points" the participants earned during Scenario 10, Gap Acceptance, which is displayed in Figure 5. The participants' baseline Druid scores correlated *positively* with the total number of good driving points [r = .31, p = .07, N = 34]. This finding is consistent with the analyses summarized in Table 3, suggesting again that a relatively poor performance on the app may have increased the participants' motivation. Similar analyses are needed for the other scenarios.

Figure 5. Total Good Driving Points, Scenario 10, Gap Acceptance (N = 36).



As shown below, a simple regression analysis showed that higher baseline DRUID scores were associated with the total number of good driving points when controlling for the participants' reported alertness after their driving simulator session.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-3.648	9.734		-.375	.710
	BASELINE	.453	.193	.395	2.349	.025
	Alertness	1.191	.660	.304	1.805	.081

a. Dependent Variable: TOTAL_GOOD

Importantly, higher DRUID baseline scores also predicted *lower* post-session alertness [$r = -.27$, $p < .001$, N = 36]. This remained the case when controlling for an index variable, "Demanding," that is the sum of the participants' four post-session ratings of how mentally and physically demanding they found DRUID and the driving simulator to be (see below). An additional regression analysis found that the baseline scores were significantly associated with lower post-session alertness when controlling for the Demanding index (result not shown).

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	4.474	1.878		2.382	.023
	BASELINE	-.095	.039	-.324	-2.469	.019
	Demanding	.123	.026	.616	4.695	<.001

a. Dependent Variable: Alertness

Conclusion

With this pilot project, we have collected a wealth of data using a novel analysis approach for driver simulator data. We have much additional work to do to mine this data, refine and expand the preliminary analyses we have reported here, and move toward manuscript preparation. Presently, we are envisioning two manuscripts: 1) a presentation of our analysis methodology, and 2) the results of the study, focusing on DRUID's validity for assessing older adults' cognitive and motor functioning and the association between baseline DRUID scores and performance on the driving simulator.

Depending on the HPL research team's interest and availability, we envision submitting a proposal to NIA to build on this pilot study. If so, our study protocol could be revised in two important ways. First, we would recruit participants who would agree to use DRUID for an extended period of time prior to their driving simulator session. We do suspect that having our study participants learn DRUID immediately beforehand made the “stereotype threat” that seniors often experience when their cognitive and motor functioning is brought into question. The revised procedure would obviate this problem while also establishing more reliable baseline scores. Second, we would develop the scenarios with our novel data analysis approach in mind to ensure that participants complete each drive and are not timed out prematurely, and to make the data extraction process less onerous. Importantly, this would give us the opportunity to reconsider how challenging the scenarios should be. As we developed the set of present mini scenarios, we were mindful of the need to avoid having the study participants experience stimulator sickness, but it may be that having only short scenarios with relatively few challenging road conditions reduced the variance in the participants’ driving performance.

As our work continues, we believe the DRUID app is a promising tool for objectively assessing whether aging senior adults should be formally evaluated to gauge their capacity to drive a car safely.