



METHOD ARTICLE

REVISED **Creating a driving profile for older adults using GPS devices and naturalistic driving methodology [version 2; referees: 2 approved, 1 approved with reservations]**

Ganesh M. Babulal¹, Cindy M. Traub², Mollie Webb², Sarah H. Stout¹, Aaron Addison², David B. Carr⁶, Brian R. Ott⁷, John C. Morris^{1,3-5}, Catherine M. Roe¹

¹Charles F. and Joanne Knight Alzheimer’s Disease Research Center, Department of Neurology, Washington University School of Medicine, St. Louis, MO, 63130, USA

²Data and GIS Services, University Libraries, Washington University in St. Louis, St. Louis, MO, 63130, USA

³Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, 63110, USA

⁴Physical Therapy, Washington University School of Medicine, St. Louis, MO, 63110, USA

⁵Occupational Therapy, Washington University School of Medicine, St. Louis, MO, 63110, USA

⁶Departments of Medicine and Neurology, Divisions of Geriatrics and Nutritional Science/Neurorehabilitation, St. Louis, MO, 63110, USA

⁷Department of Neurology, Warren Alpert Medical School, Brown University, and Rhode Island Hospital, Providence, Rhode Island, USA

v2 **First published:** 26 Sep 2016, 5:2376 (doi: [10.12688/f1000research.9608.1](https://doi.org/10.12688/f1000research.9608.1))
Latest published: 07 Dec 2016, 5:2376 (doi: [10.12688/f1000research.9608.2](https://doi.org/10.12688/f1000research.9608.2))

Abstract

Background/Objectives: Road tests and driving simulators are most commonly used in research studies and clinical evaluations of older drivers. Our objective was to describe the process and associated challenges in adapting an existing, commercial, off-the-shelf (COTS), in-vehicle device for naturalistic, longitudinal research to better understand daily driving behavior in older drivers.

Design: The Azuga G2 Tracking Device™ was installed in each participant’s vehicle, and we collected data over 5 months (speed, latitude/longitude) every 30-seconds when the vehicle was driven.

Setting: The Knight Alzheimer’s Disease Research Center at Washington University School of Medicine.

Participants: Five individuals enrolled in a larger, longitudinal study assessing preclinical Alzheimer disease and driving performance. Participants were aged 65+ years and had normal cognition.

Measurements: Spatial components included Primary Location(s), Driving Areas, Mean Centers and Unique Destinations. Temporal components included number of trips taken during different times of the day. Behavioral components included number of hard braking, speeding and sudden acceleration events.

Methods: Individual 30-second observations, each comprising one breadcrumb, and trip-level data were collected and analyzed in R and ArcGIS.

Results: Primary locations were confirmed to be 100% accurate when compared to known addresses. Based on the locations of the breadcrumbs, we were able to successfully identify frequently visited locations and general travel patterns. Based on the reported time from the breadcrumbs, we could assess number of trips driven in daylight vs. night. Data on additional events while driving allowed us to compute the number of adverse driving alerts over the course of the 5-month period.

Open Peer Review

Referee Status: ? ✓ ✓

	Invited Referees		
	1	2	3
REVISED			
version 2 published 07 Dec 2016			
version 1 published 26 Sep 2016	? report	✓ report	✓ report
1 Frank-Dietrich Knoefel , Bruyère Research Institute Canada			
2 Xueqin Qian , Department of Educational Psychology USA			
3 Monique M. Williams , VITAS Healthcare USA			

Discuss this article

Comments (0)

Conclusions: Compared to cameras and highly instrumented vehicle in other naturalistic studies, the compact COTS device was quickly installed and transmitted high volumes of data. Driving Profiles for older adults can be created and compared month-to-month or year-to-year, allowing researchers to identify changes in driving patterns that are unavailable in controlled conditions.

Corresponding authors: Ganesh M. Babulal (babulalg@neuro.wustl.edu), Catherine M. Roe (cathyr@wustl.edu)

How to cite this article: Babulal GM, Traub CM, Webb M *et al.* **Creating a driving profile for older adults using GPS devices and naturalistic driving methodology [version 2; referees: 2 approved, 1 approved with reservations]** *F1000Research* 2016, 5:2376 (doi: [10.12688/f1000research.9608.2](https://doi.org/10.12688/f1000research.9608.2))

Copyright: © 2016 Babulal GM *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: National Institute on Aging [R01AG043434, R01AG43434-03S1, P50-AG05681, P01-AG03991, P01-AG026276]; Fred Simmons and Olga Mohan, and the Charles and Joanne Knight Alzheimer's Research Initiative of the Washington University Knight Alzheimer's Disease Research Center.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: Ganesh Babulal, Cindy Traub, Molly Webb, Sarah Stout and Aaron Addison declare they have no competing interest. Brian Ott: Grants and funds: Eli Lilly, Avid, Roche, TauRX, Merck, Univita, NIH/NIA; Honoria: NHTSA: Medscape; Consultant: Accera (DSMB). David Carr: Grants and funds: Missouri Department of Transportation; Honorarium: Harvard Speaker; Consultant: Traffic Injury Research Foundation, Advanced Drivers Education Products and Training; Medscape. John Morris: Grants/Funds: Healthy Aging and Senile Dementia, Antecedent Biomarkers for AD: The Adult Children Study, The Dominantly Inherited Alzheimer Network, and Alzheimer Disease Research Center. Honoraria: Cherkin Lecture, Chinese Society, 13th Eibsee Meeting (Keynote Speaker), Korean Dementia Association, and DZNE research center Magdeburg Symposium. Consultant: Lilly USA; ISIS Pharmaceuticals; Charles Dana Foundation. Royalties: Blackwell Medical Publishers; Taylor & Francis. Board Member: Board of Directors American Academy of Neurology (AAN). Catherine Roe: Grants/Funds: NIH/NIA.

First published: 26 Sep 2016, 5:2376 (doi: [10.12688/f1000research.9608.1](https://doi.org/10.12688/f1000research.9608.1))

REVISED Amendments from Version 1

Information comparing the GPDAS device to other naturalistic driving methodologies and their associated references have been added into the Discussion section. Specific definitions of speeding, hard braking and hard acceleration pertaining to the device have also been added. Per reviewer suggestions, other clarification has been added to the Background, Methods and Results sections.

See referee reports

Background

The high risk of crashes and decline in driving ability among older adults with Alzheimer disease is well documented; the risk of injury and mortality from motor vehicle crashes increases with age and dementia severity¹. Our research program seeks to understand driving behavior among older adults, particularly as it occurs on a day-to-day basis as people travel in their own environments. However, evaluation of driving behavior in older adults largely occurs with methodologies that use controlled conditions such as on-the-road tests and driving simulators, and to a lesser extent, self-report and diaries²⁻⁵. To better meet our research needs, we explored newer methodologies to study naturalistic driving behavior longitudinally, in a cost effective and unobtrusive manner⁶.

Recent technological advances in global positioning systems (GPS) and geographic information systems (GIS) techniques allow evaluation of driving behavior in the actual environments in which individuals drive⁷. Newer in-vehicle GPS devices are unobtrusive and typically provide data on date, time, speed, longitude and latitude regarding where a vehicle is driven^{8,9}. In-vehicle GPS/GIS devices are an emerging methodology employed to better understand driving *in situ* and compare differences between driver self-report and GPS data obtained from a vehicle¹⁰. As a result, naturalistic driving research employing this methodology seeks to understand driving behavior by analyzing continuous, objective data collected by in-vehicle devices to determine patterns and the influence of personal, temporal and environmental factors^{8,11}.

The evolving field of naturalistic driving and the proliferation of custom and commercial off the shelf (COTS) in-vehicle devices have resulted in numerous different outcomes and GIS analytical techniques^{9,12}. However, some challenges accompany GPS data use, including extensive post-processing of large volumes of data, variability with temporal and spatial aspects of the data, and higher cost associated with the technology and data collection. Consequently, the monitoring periods in some recent studies using GPS and GIS are limited to capturing data for analysis from a timespan ranging from weeks to 2 months^{10,12}. However these short periods may be too brief to capture relevant driving behaviors.

To more accurately monitor key driving naturalistic driving behaviors, we piloted a new methodology adapting a COTS in-vehicle device to study naturalistic driving behavior longitudinally, in a cost effective and unobtrusive manner. Our objective for this pilot is to describe methodological challenges associated with

adapting a COTS in-vehicle device that captures and synthesizes GPS data for processing and analysis using GIS techniques. We also quantify spatial and temporal patterns associated with driving behavior to construct driver profiles to evaluate how driving behavior changes longitudinally.

Methods

Participant data. Data were collected from participants enrolled in a longitudinal study assessing preclinical Alzheimer's disease and driving performance (R01 AG043434) at Washington University School of Medicine in St. Louis. Participants had normal cognition, were 65 years or older, had a valid driver's license, drove at least once per week in a non-adapted vehicle, met minimal visual acuity for state requirements, and had Alzheimer's disease biomarkers (cerebrospinal fluid or brain imaging) objectively measured and available within the last two years. All study protocols, consent documents and questionnaires were approved by Washington University Human Research Protection Office.

Data collection and processing. We used the COTS Azuga G2 Tracking Device™ (Model 850: Azuga Inc, San Jose, California), which we refer to as a global positioning data acquisition system (GPDAS). The GPDAS plugs into the on-board diagnostic system port (OBDII) and is powered by the vehicle's battery. Installation requirements limit vehicles to those manufactured in 1996 or later since earlier years were not equipped with an OBDII port. Data (vehicle speed, latitude, longitude) were collected from the moment ignition was turned on and until it was turned off, with a collection interval set at every 30 seconds. Individual 30-second observations are referred to as a "breadcrumb". Location data were also collected every three hours when the ignition was off. Additionally, aggressive driving incidents such as hard braking, speeding and sudden acceleration were recorded in the trip log. Data were collected and simultaneously transmitted via Bluetooth Low Energy to secured servers. On a daily basis, the data were aggregated by Azuga and made available for download via secured servers.

Two distinct file types available from Azuga were used in our analysis – Breadcrumb files and Activity files. Within the daily Breadcrumb comma separated values (csv) file, each row consisted of one observation ("breadcrumb"), typically at a 30 second interval for a specific vehicle at an instant of time. Each breadcrumb identified the vehicle by a 10-digit code and additionally reports latitude, longitude, vehicle speed, nearest address (reverse geocoded by Azuga), coordinated universal time (UTC) and date, odometer reading, and event type. The event type field identified whether the given breadcrumb was associated with a regular observation or special event such as ignition on/off or aggressive driving. The event type field could also contain codes indicating specific issues such as a low battery level in the vehicle, a connection or disconnection of the device, or a malfunction in the device hardware. Additional fields gave data about the peak speed and average speed of an over-speeding event, as well as initial and final speeds of braking or acceleration events characterized by a rapid change in the vehicle's velocity.

The second file type received from Azuga was the daily Activity csv file. Each row in the Activity file represented one trip taken by

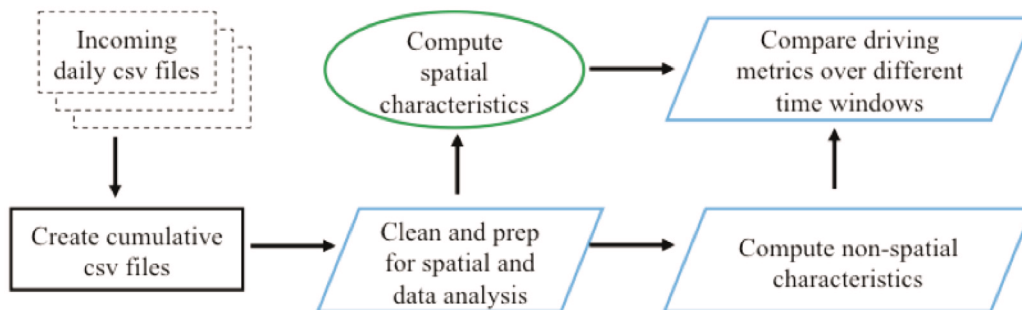
a single vehicle. Available observations about each trip included the date and start time (in UTC), the starting and ending locations (latitude, longitude, and reverse-geocoded address), the duration/length of the trip in seconds and in distance (rounded to the nearest tenth of a kilometer, then reported in miles), the average and maximum vehicle speed, and the number/duration of aggressive driving events such as sudden acceleration, hard braking, and over-speeding. Preliminary data processing used a Powershell script to compare headers from the incoming Breadcrumb and Activity files to ensure the structure was consistent, and then combined the daily files from the time period of interest into two large comprehensive csv files (one each for Breadcrumbs and for trip-level Activity). These two large csv files were read into the statistical analysis program R as data tables for further analyses. For the remainder of the manuscript, the term *breadcrumb* refers to a single observation of one vehicle at a specific location and single moment in time, while a *trip* represents a set of locations (breadcrumbs) occurring between the ignition on and ignition off of a specific vehicle. Over the first five months, over 400,000 breadcrumbs representing approximately 12,000 trips were collected for the 20 vehicles.

Initial processing steps taken in R examined the condition of the incoming data for errors and anomalies, then created additional fields for use in aggregating the data, as well as the spatial processing stages. Since all times were reported in UTC and our participants were in the continental United States, time zone calculations were performed to accurately transform the incoming timestamp to local time. Many points in the Central Standard Time Zone were classified as such within R using a bounding rectangle with maximum/minimum latitude and longitude encapsulating the majority of the Central Standard Time Zone. For points close to the boundary of time zones, GIS was used to determine the appropriate zone. This was done by comparing the breadcrumb location against a set of polygons representing the extent of each time zone to determine in which time zone polygon the breadcrumb location fell in. Local time was needed to understand driving activity or avoidance during specific times of day (rush hour, daylight, etc.). The R package lubridate was used to convert UTC time to local time, while

the R package RAtmosphere allowed for computations of sunrise and sunset at a given latitude/longitude. These computations were added as additional columns in the data tables. A summary of the workflow is given in Figure 1.

To clean the incoming data and prepare these for spatial processing, data were checked to ensure that two criteria were met: (1) each observation occurred within the continental United States and (2) no two observations for the same vehicle had identical timestamps. Certain device actions (being connected, disconnected, or plugged into a different vehicle) caused the GPDAS to report latitude and longitude values of 0, or in one case, those of a location in Egypt. Additionally, for some vehicles that started a trip immediately after plugging in the GPDAS, the time delay required to connect to a sufficient number of GPS satellites to register locational data caused a sequence of observations with latitude and longitude equal to 0. Due to uncertainty about the location of the vehicle at times where the latitude or longitude was reported outside the continental United States, those trips, including associated breadcrumbs were removed from analyses. The number of breadcrumbs impacted was less than 1.6% of all incoming breadcrumbs, with the vast majority of Figure 1 (6392 out of 6529) representing one vehicle whose GPDAS had a malfunction causing no locational data to be collected for multiple weeks. Removing the vehicle with a faulty GPDAS from the computation reduced the number of breadcrumbs removed by the first criteria to 137, less than 0.04% of the total number of breadcrumbs collected. The second criteria removed 12 breadcrumbs that were exact duplicates of other breadcrumbs.

Further data cleaning was required to compile a set of complete trips taken by each driver. Trip-level data were accessible in two ways from the incoming data stream. The Activity files contained summary information about the start, end, and length of each trip, while the Breadcrumb files offered a finer level of locational detail within the trip. Approximately 1.6% (n=203) of the incoming activity records contained NA values as latitude and longitude of the trip end. Typically this was caused by either a loss of GPS signal



The shapes in the workflow represent the software used at each stage: a black rectangle for Powershell, a blue parallelogram for R, and a green oval for ArcGIS/ArcPy.

Figure 1. Data workflow required to generate driving metrics over different time scales.

(such as parking in an underground structure) or a peculiarity of the incoming activity data stream, in which a second recorded trip start occurred several seconds after the first, which was then “abandoned” as a meaningful trip in the data stream. An additional 1.8% of reported trips (n=229) contained a value of 0 for the starting latitude or longitude. Most of these (217) were from the aforementioned known defective device that transmitted large numbers of zeros within the breadcrumb data. These were marked for removal.

Analysis. Data analysis and management for spatial operations in GIS, used ArcGIS 10.3.1 and the ArcPy Python site package (Environmental Systems Research Institute, Redlands, CA, USA). Spatial data were stored as feature classes in file geodatabase format. Time zone computations were exported from ArcGIS as a csv file and merged back in with the data table in R for further computations.

Spatial analysis. Using the latitude and longitude for each breadcrumb, point feature classes were created for each driver by exporting the results of the Make XY Event geoprocessing operation. These point feature classes served as the basis for all subsequent spatial analysis.

Road analysis. To determine the characteristics of the road over which the participant was traveling at the time of breadcrumb

recording, proximity analysis was performed on each breadcrumb relative to a street centerline dataset. The Near geoprocessing operation was used to identify the street centerline feature closest to the street feature for each breadcrumb. The output of the Near geoprocessing operation is the addition of two attributes to the breadcrumb feature class. These attributes are NEARFID, the unique identifier of the nearest street feature, and NEARDIST, the distance from the target breadcrumb to the nearest street feature. The NEARFID value was used to retrieve attributes of the street feature nearest to the breadcrumb, such as the road name, Census Feature Class Code (CFCC), road type, and average speed (proxy for speed limit). **Figure 2** shows a sample of breadcrumbs and their proximity to the street centerline features. Attributed values from the nearest street feature were applied to each breadcrumb using a series of Cursors. Cursors are iterator tools available in the ArcPy code library that can read, update and create features in existing spatial datasets (ArcGIS Help 2015).

Driving Areas. Driving Area was defined as the smallest polygon that encompassed all breadcrumbs for a driver during a given time period. The Minimum Bounding Geometry geoprocessing operation was used to produce convex hull polygons representing the weekly and monthly Driving Areas for each driver.

Mean Center. The Mean Center was defined as the geographic center of all breadcrumbs for a driver during a given time period. The Mean Center geoprocessing operation was used to produce points representing the weekly and monthly Mean Center for each driver. The operation was based on spatial location of the breadcrumbs only and was not weighted by any attribute.

Primary Locations. The participants’ most commonly-visited locations (Primary Locations) were identified in order to perform spatial analysis on aspects of the drivers’ behavior relative to familiar areas. The participant’s home and/or workplace were assumed to be the most frequent origin or destination of the majority of the trips recorded by the GPDAS. It was crucial that these locations be identified in a dynamic and automated way to achieve scalability of the data processing workflow. A visual examination of the data for a small sample of participants showed that an often-visited location could appear as a dense cluster of breadcrumbs. It was assumed that the densest cluster, or the cluster with the most ignition on breadcrumbs, would be the Primary Location. Clusters of ignition on breadcrumbs were identified using the Aggregate Points geoprocessing operation. The Aggregate Distance parameter was set to 20 feet after visually locating and measuring ignition on breadcrumb clusters on a small sample of participants. The output of the Aggregate Points operation was polygon feature class with features encompassing clusters of three or more points within the Aggregate Distance parameter value. The breadcrumbs located within each polygon were counted and compared to the total number of ignition on breadcrumbs for the participant to determine if the polygon represented a Primary Location. The Feature To Polygon geoprocessing operation was used to produce a point feature at the centroid of each Primary Location polygon, thus providing a single point that was used as the Primary Location in further analyses (**Figure 3**).

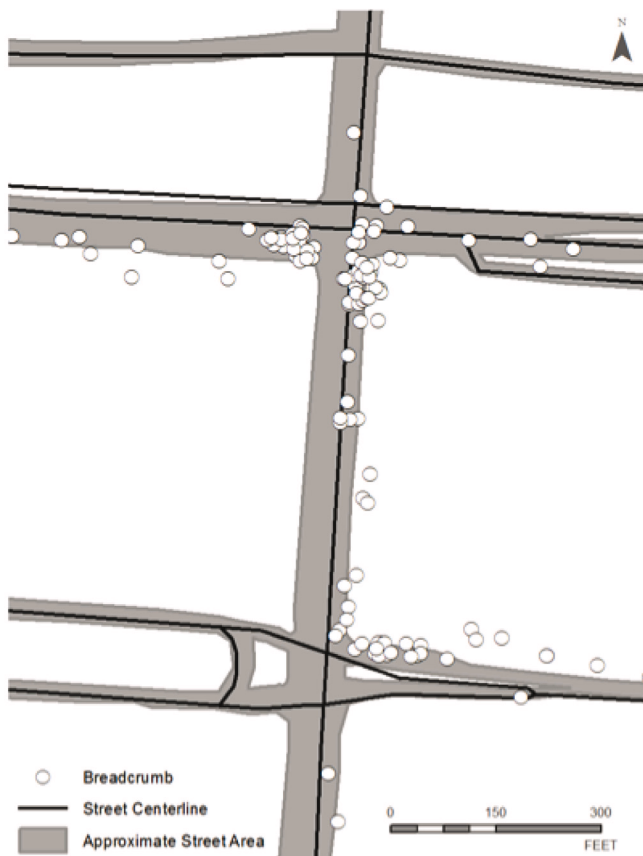


Figure 2. Breadcrumbs and street centerline features used in near analysis.

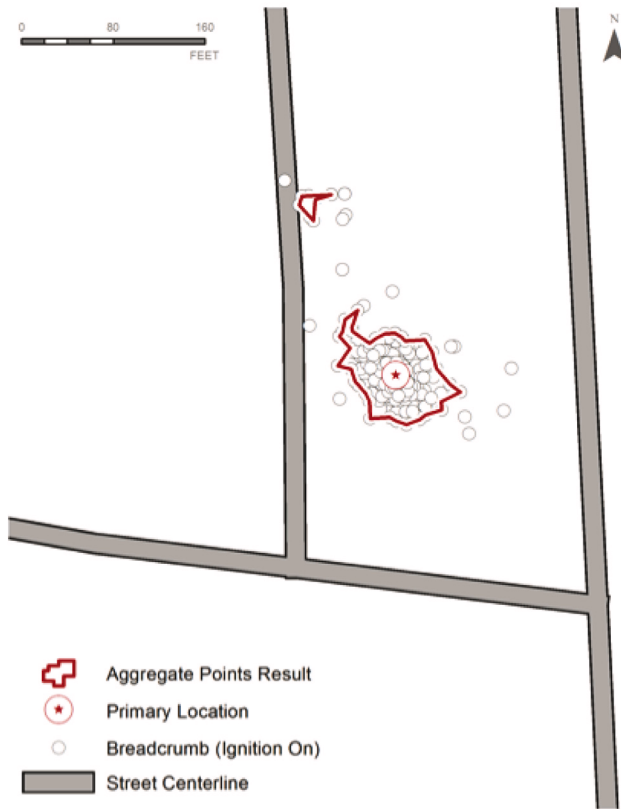


Figure 3. A primary location polygon created using the aggregate points geoprocessing operation with a primary location point placed at its centroid.

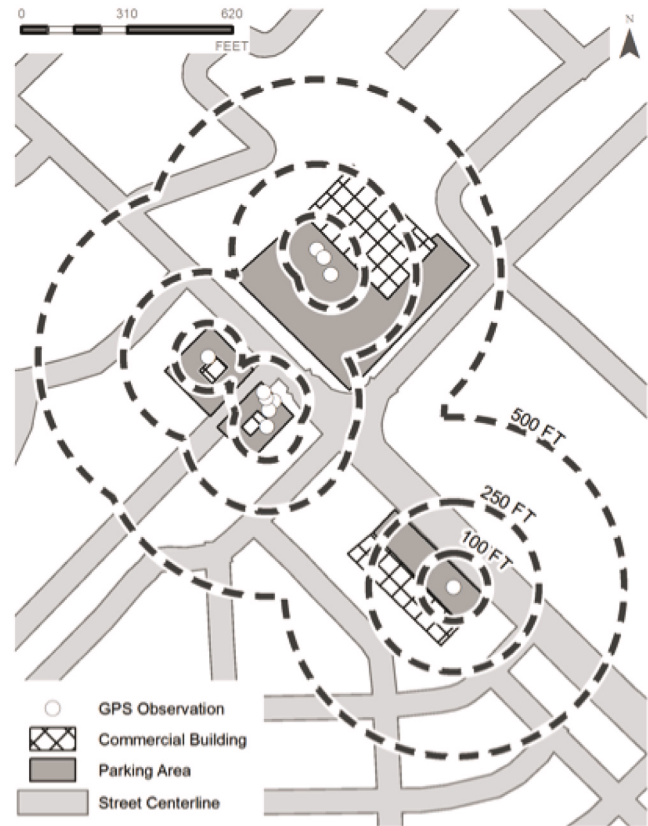


Figure 4. Unique destination sample with multiple buffers.

Unique destinations. Unique destinations are defined as separate locations visited by participants during a given timeframe. The Buffer geoprocessing operation was used to create circular polygons with radii of 100, 250 and 500 feet around each breadcrumb indicating an ignition on event. The varying buffer operations were performed to establish a threshold at which two or more distinct breadcrumbs occurring within the same radius during the same time period would be combined as the same destination. For example, a participant who visited a shopping center twice in the same month may park at opposite ends of the large parking area for each separate visit. However, this shopping center should be counted as a single destination for the target time period.

The Dissolve geoprocessing operation was used to merge the circular polygons so breadcrumbs within the three distance thresholds would be counted as a single destination. Figure 4 shows a sample of ignition on breadcrumbs in a selected area during a single month. The groups of breadcrumbs within close proximity to each of the commercial buildings occur on different days within the same month. The 500 foot buffer polygon encompassed all four separate commercial destinations and would be counted as a single destination for that month. The 250 foot buffer would combine the northernmost commercial area with the two areas

to the southwest, creating a single destination from three distinct destinations. The 100 foot buffer separated the three separate destinations into two destinations, combining only the two smallest commercial areas into a single destination (Figure 4).

Results

Comprehensive driver profiles. A breadcrumb is one data point in time (at 30-second interval) that contains location, time, date and speed of a vehicle. A single trip could have hundreds of breadcrumbs that are aggregated and over time can provide specific information about driving patterns and behaviors. The steps discussed in the methodology section resulted in the creation of a driving profile for each driver that could be examined over the course of a study. Driver profiles included spatial, temporal and behavioral components. Spatial components included Primary Location(s), Driving Areas, Mean Centers and Unique Destinations. Temporal components included number of trips taken during different times of day. Behavioral components included number of hard braking, speeding and sudden acceleration events. Across the five months, the mean and standard deviation for the total number of trips was 552.7 (209.5) and for average miles per trip was 6.8 (3.2).

Primary Locations. A driver’s Primary Location was designated as the location that encompassed at least 10 percent of the driver’s Ignition On breadcrumbs. Since participants are over the age of 65, in most cases, participants had a single Primary Location, assumed to be their home/residence, though some participant results showed two Primary Locations. In most cases, the count for the cluster polygon with the highest count of breadcrumbs was significantly higher than the counts for the other two polygons. The exception is Participant C, where two cluster polygons have breadcrumb counts over 10 percent of the total breadcrumbs (Figure 5). Participant C has two Primary Locations based on the percentage of driver’s ignition on events. Primary Locations were compared against the known addresses from the participants and confirmed to be 100 percent accurate, including participant C who is known to have two homes.

Driving Area and Mean Center. The Driving Area polygons resulting from the methodology varied based on the extent of the breadcrumbs for each driver over time. Analysis showed that a participant’s driving areas could often have large portions of overlap from week to week or month to month. Mean Centers were expected to be clustered around the participant’s Primary Location. However, this was not the case when participants had more than one Primary Location. See Participant C in Figure 5. Participant C had two designated Primary Locations and as a result, the Mean

Centers for this participant tend to be located between the two Primary Locations. The combined Driving Area polygons, Mean Centers and Primary Locations make up the spatial profile for study participants. Spatial profiles for a sample of participants are visualized in Figure 5. Each grey polygon represents the driving area for a single month for each participant. Monthly Mean Centers are represented with white boxes and red stars indicate the Primary Location for each participant.

Driving Areas can vary greatly month to month for some participants, while other participants tend to have little monthly variation in their driving area. The monthly Driving Area polygons for participants C and D show large portions of overlap, while participants A, B and E show large portions of Driving Area unique to a single month timeframe. Common Driving Area can be quantified by calculating the overlapping area from month to month and overall overlapping area for the five-month study period. The month to month variation in overlapping Driving Area is shown in Figure 6 and reinforces the large amount of overlap from month to month for participants C (first line from top) and D (second line from top).

The ratio of overlapping Driving Area over total Driving Area examines the relationship between commonly driven routes and total driving space. In Figure 7, participant C shows little variation in monthly driving area during the study timeframe with over 70%

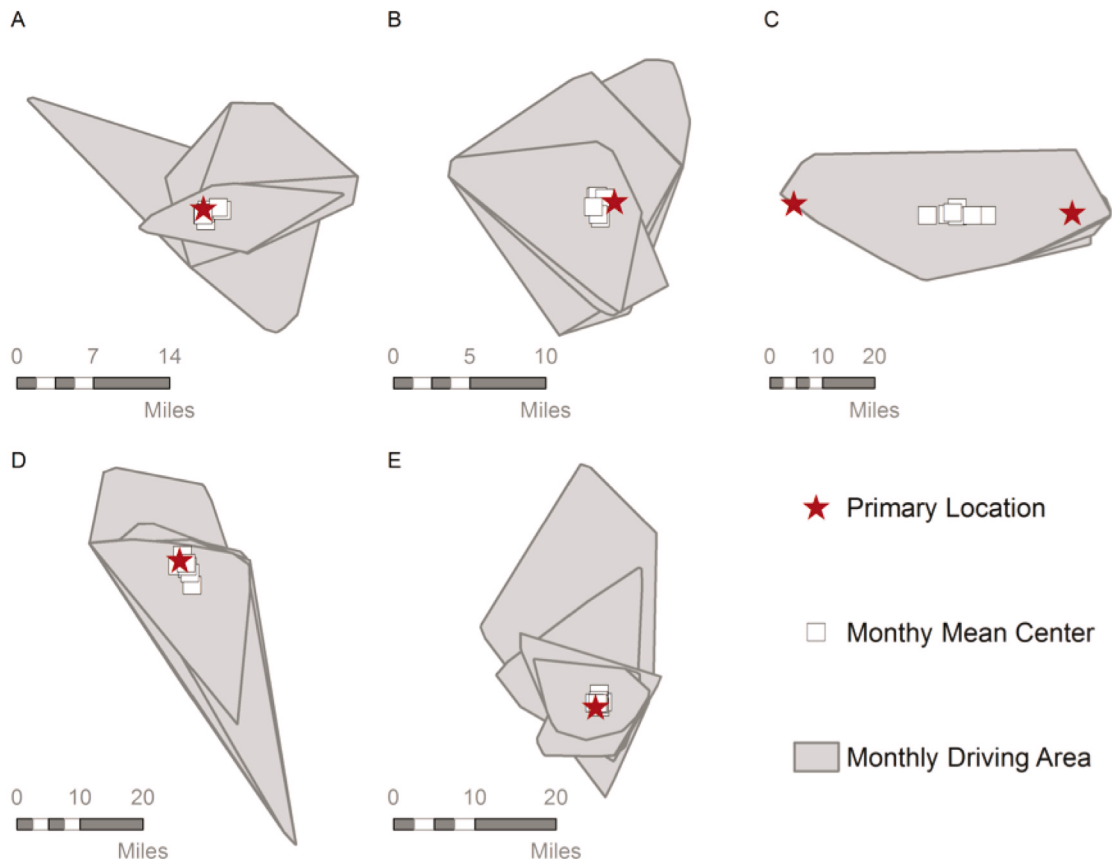


Figure 5. Spatial profiles.

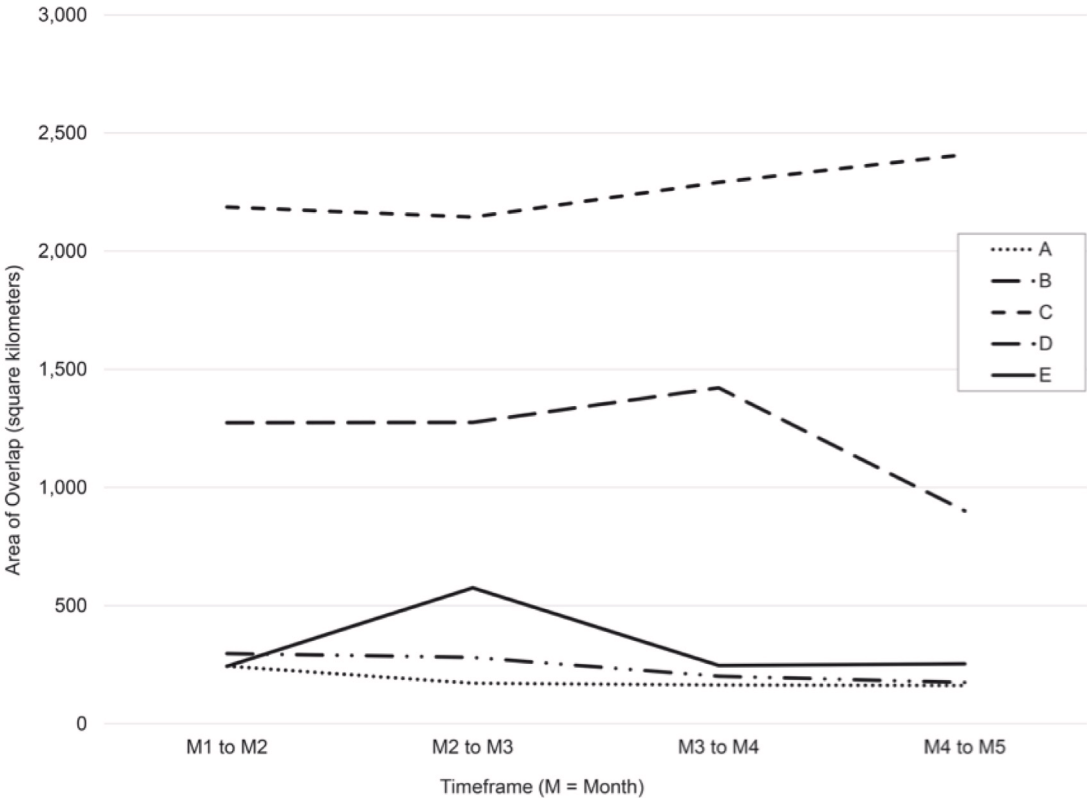


Figure 6. Area of overlap between Driving Areas from month to month.

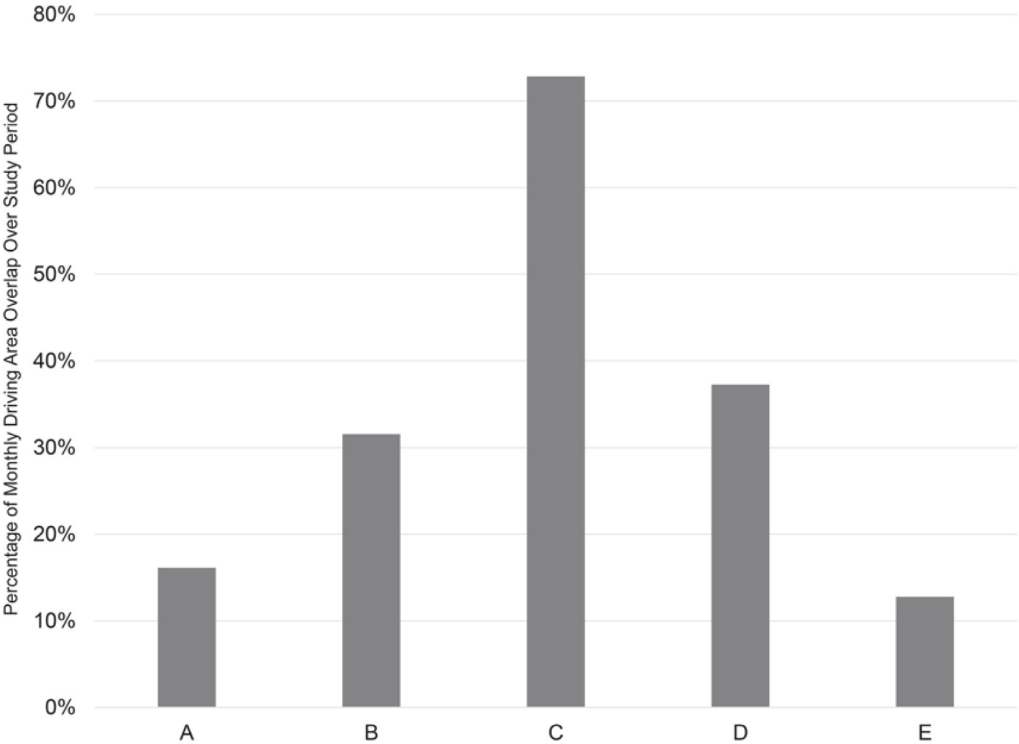


Figure 7. Ratio of overlapping driving area to total driving area over the 5 month period.

of the total Driving Area being common to all months. Participant E shows the least amount of overlapping area with less than 15% of the total Driving Area being common to all months.

Unique Destination. The results of performing the Unique Destinations methodology showed varying results by driver. While some drivers showed similar counts of Unique Destinations each month, other drivers showed counts of Unique Destinations that varied greatly from month to month (Figure 8). In most months for many drivers, the counts of Unique Destinations derived by using the 100, 250 and 500 feet buffers varied by buffer size. However, if a driver’s destinations were particularly spread out, the buffer size was less consequential. Overall, the results show that the 100 feet buffer should be used to obtain the most accurate count of unique destinations for the participants within each time period.

Trips driven in daylight vs. night-time. The results of the number of trips driven during the day vs. night showed variation across individuals and intra-individual change across different months. Figure 9 displays the number of trips driven during day and night for five participants from July (7) to November (11). Night driving is associated with a three times greater risk of traffic death and increased fatigue and perceived danger^{13,14}. The majority of trips driven by 4/5 participants were driven during the day. For participants B and D, the number of trips generally declined from month 1 to month 5 without a significant change in their number of trips driven at night. However, Participant C had a higher total number of trips for months 4–5 and increased night driving

compared to months 5–7. Participant A reduced their night driving and total number of trips taken from month 3 to 5 while participant B showed little change in night driving behavior. Given that the time window of our study represents months when the hours of daylight available are steadily decreasing, the decrease in total number of trips combined with the lack of a corresponding increase in trips taken at night may suggest that the driver A in our study made deliberate adjustments to avoid night-time driving. Finally, more trips were started during dusk, compared to dawn.

Adverse driving behavior. Three alerts (speeding, hard braking, and hard acceleration) were identified by the GPDAS that reflect adverse driving behavior independent of the environmental driving context. Speeding was defined as driving more than ≥ 6 miles per hour above posted speed limit in an area. Duration and maximum speed were gathered and available from the device, however, only the counts of alert were reported here. Hard braking and hard acceleration were defined as a change (decrease/increase) in speed of 8 to 10 miles per hour in one second, respectively. Counts were also reported for hard braking and hard acceleration. All definitions were adopted from the insurance industry by the device vendor. The GPDAS does not capture data on traffic flow or congestion, weather patterns, inclement conditions, or other factors (e.g. altered mental state) that may impact the driver behavior. Figure 10 presents data on hard braking, sudden acceleration and speeding for five participants across the five months. Similar to the spatial and temporal analyses, there was a wide variation among

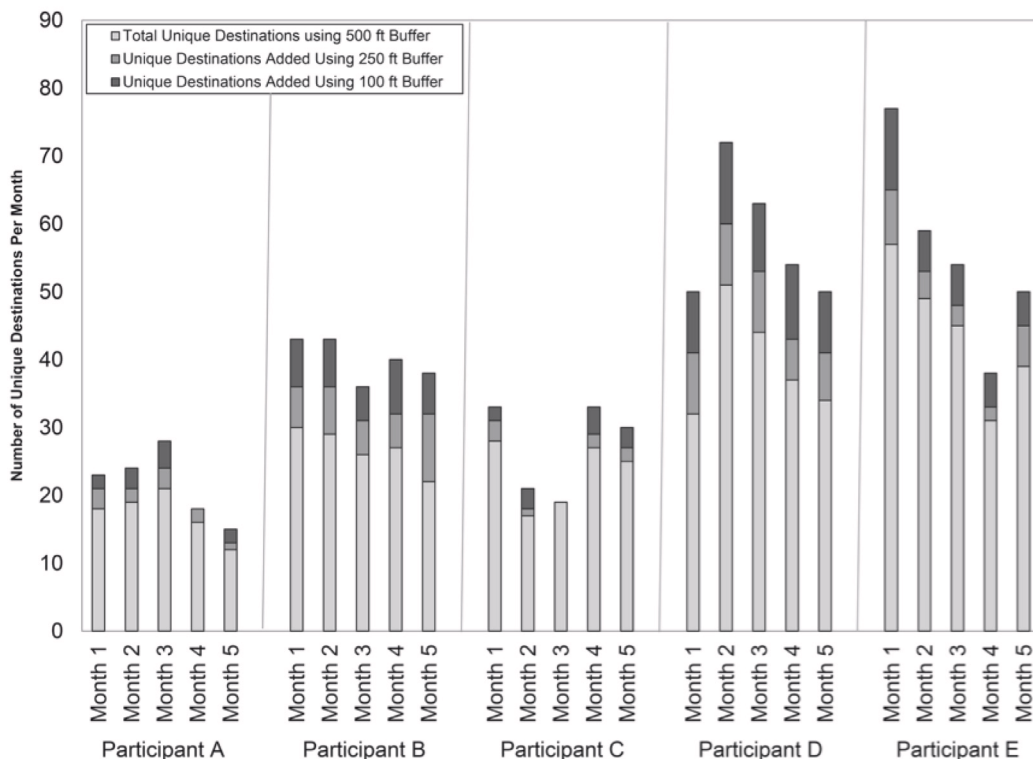


Figure 8. Counts of unique destinations for 5 participants.

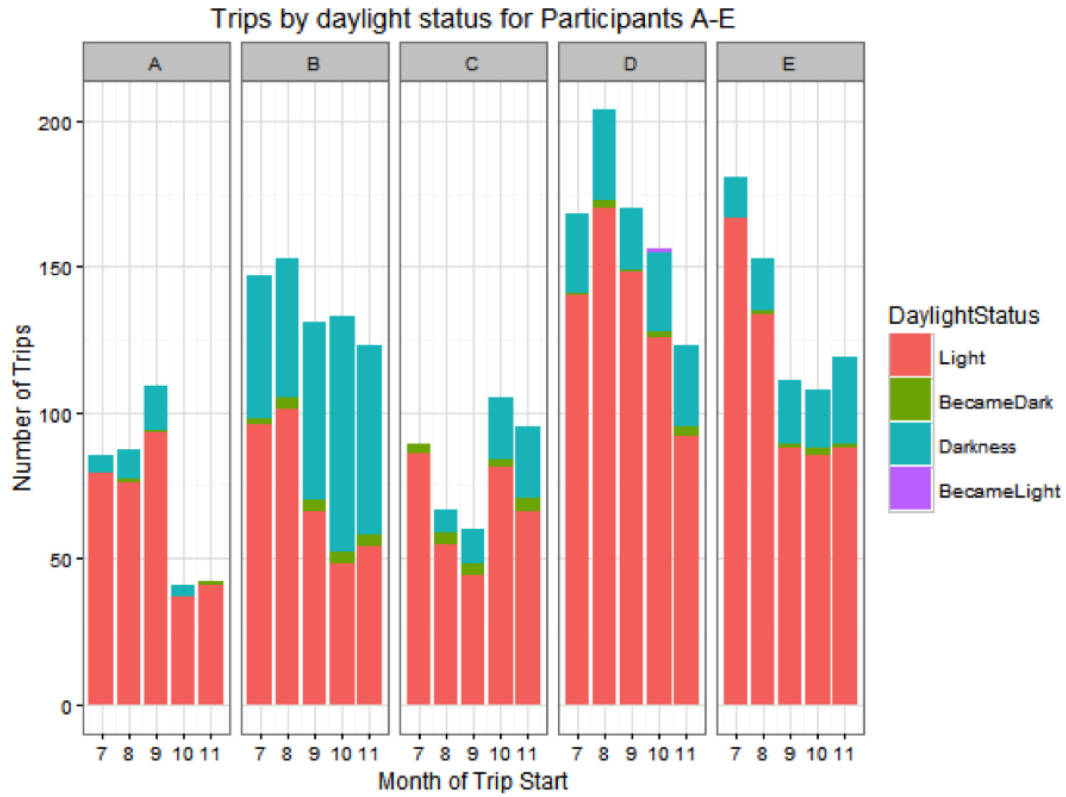


Figure 9. Number of trips driven during day and night.

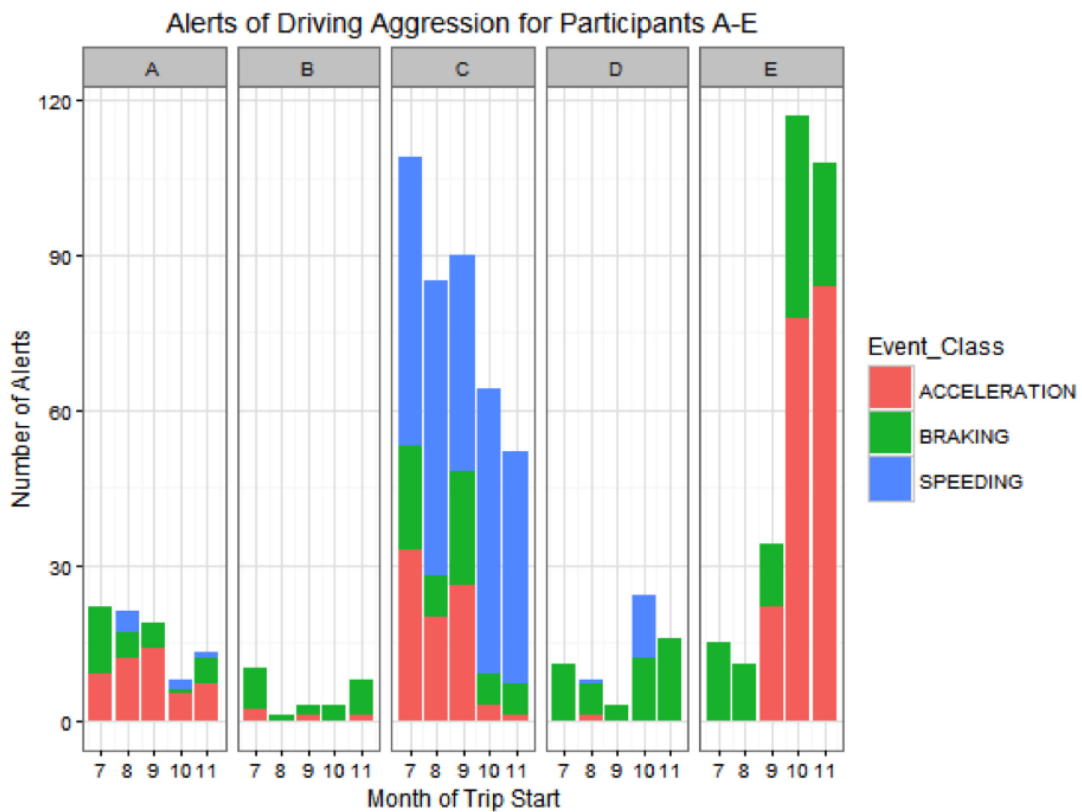


Figure 10. Total number of alerts across the 5 months.

the participants. The difference between the least and most aggressive drivers shown here is dramatic: Participant B recorded 25 total alerts while Participant C recorded 400. Participants B and D had no speeding alerts, while participant C recorded all three types of aggressive driving patterns in all 5 months. Participant D recorded three times as many braking events as speeding and sudden acceleration events combined. In months 10 and 11, Participants D and E show a marked increase in aggressive habits, while Participants A and C seem to decline in aggression. The inter-individual variation in driving alerts over the five months may be a reflection of driver preference or style, the driving environment or the interaction between both. While the data presented in [Figure 10](#) is a total count of alerts, it is possible to examine the frequency of trips containing one or more alerts. It is unlikely that the high number of alerts among some participants (*e.g.* C, E) may be solely attributed to the driving environment.

In summary, the COTS GPDAS device was able to capture objective driving behavior. Data obtained provided a foundation for creating a Naturalistic Driving Profile that included spatial, temporal and behavioral components. This methodology allows us to track a number of variables describing the driving behaviors and patterns of participants over time. We were able to confirm the accuracy of the methodology of identifying the Primary Locations by comparing the results to the actual addresses reported by the participants. Based on the locations of the breadcrumbs, we were able to successfully identify frequently visited locations and general travel patterns. Based on the reported time from the breadcrumbs, we could assess number of trips driven in daylight vs. night-time. Data capturing special events allowed us to compute the number of adverse driving alerts over the 5-month period.

Discussion

This pilot study presented the feasibility of adapting a COTS GPS device to examine daily driving behavior and associated changes in a cohort of cognitively-normal older adults. The ability to understand changes in driving behavior in the actual environments people drive has been unavailable until recently.

The GPDAS provided continuous driving data that was used to develop a unique Naturalistic Driving Profile combining spatial, temporal and behavioral aspects of driving. Specifically, we were able to obtain date, time, location and a set of metrics that balanced the ability to measure consistency and change in driving behavior, without over-collecting data or over-burdening research participants. The complexities and obstacles of working with large datasets have been well documented¹⁴. The key methodological challenges in this research included: 1) synchronizing data collection from the GPDAS and the vendor servers, 2) efficiently processing and error checking the ‘big data’ on a daily basis 3) developing data cleaning procedures for common errors (*e.g.* device removal or signal loss) and uncommon errors (*e.g.* device failure) and 4) synthesizing the data for management and analyses in R and ArcGIS.

This naturalistic driving methodology provides several advantages to understanding driving behavior over conventional methodologies. The GPDAS can be used to simultaneously monitor real-time driving behavior in a large cohort across the continental United States. The GPDAS’ great strength comes in being able to observe

individuals and compare intra-individual change over a long period of time. Additionally, the ease of installation (less than 1 minute), no vehicle modification, minimal effort from participants and seamless data acquisition and transmission strengthens its utility.

Given this was a COTS GPS device, there are some limitations with using this device and methodology. Recent naturalistic driving studies have implemented and used more sophisticated systems including cameras, installed hard disk drives and a combination of sensors hardwired to a vehicle’s engine and wireless sensors^{16–18}. While we sampled data every 30 seconds, other studies have sampled data every 1 to 5 seconds to increase the fidelity and robustness of their results¹⁹. The length of this pilot was shorter when compared to other studies (*e.g.* CanDrive) that have published driving data from several years of data collection²⁰. Some studies have used radio frequency identification tags to identify the driver and anonymize home locations of their participants²¹. Driver identification was limited to participant self-report. The vendor now offers a Bluetooth Low Energy (BLE) beacon the size and weight of a credit card that may be placed in a wallet or purse. The BLE beacon automatically pairs with the GPDAS when the participant is in the driver seat to identify the driver. This simple solution is automatic, requires no participant effort, conveniently syncs with the devices data stream and is downloaded with the device’s data. We were not able to detect under-speeding as robustly as we had hoped, due to a variety of confounding factors such as traffic, construction speed limit changes, and the granularity of the breadcrumbs. The unknown product life of GPDAS devices poses a particular challenge for longitudinal naturalistic driving studies. We detected several potential warning signs for device failure, and were able to take proactive steps to order replacement devices when failures were detected. However, such replacement is not simple for some study participants since it requires travel to our facility for a new device and may reduce their willingness to remain in the study. Finally, it is important to consider the goals, outcomes and amount of participant burden when selecting a methodology for longitudinal studies assessing driving performance and behavior.

Ethical approval and consent to participate: All participants were recruited and tested at Washington University School of Medicine. Written informed consent to use and publish clinical details was obtained from all participants. All aspects of the study were approved by the Washington University Institutional Review Board.

Dataset 1. Dataset: Creating a driving profile for older adults using GPS devices and naturalistic driving methodology

<http://dx.doi.org/10.5256/f1000research.9608.d135843>

Dataset legend: id: Participant ID; numofdays: Number of days driven; numtripsover5mo: Number of trips driven over 5 months; trips_at_night5mo: Number of trips driven at night over 5 months; trips_w_HB5mo: Number of trips driven with hard braking events over 5 months; trips_w_SA5mo: Number of trips driven sudden acceleration events over 5 months; trips_speeding5mo: Number of trips driven with speeding events over 5 months; hours_speeding5mo: Number of hours speeding over 5 months; tot_dist_driven5mo: Total distance (miles) driven over 5 months; tot_drv_hrs5mo: Total number of hours driven over 5 months; avg_trip_miles5mo: Average trip miles over 5 months; avg_trip_mins5mo: Average trips minutes over 5 months; trips_at_night: Number of trips at night over 5 months.

Data availability

F1000Research: Dataset 1. Dataset: Creating a driving profile for older adults using GPS devices and naturalistic driving methodology, [10.5256/f1000research.9608.d135843](https://doi.org/10.5256/f1000research.9608.d135843)²²

Author contributions

GMB: study concept and design, acquisition of subjects and data, analysis and interpretation of data, and preparation of manuscript; CMT: acquisition data, analysis and interpretation of data, and preparation of manuscript; MW: acquisition data, analysis and interpretation of data, and preparation of manuscript; SHS: analysis and interpretation of data, and preparation of manuscript; AA: study concept and design and preparation of manuscript; DBC: interpretation of data, and preparation of manuscript; BRO: interpretation of data, and preparation of manuscript; JCM: interpretation of data, and preparation of manuscript; CMR: study concept and design, analysis and interpretation of data, and preparation of manuscript.

Competing interests

Ganesh Babulal, Cindy Traub, Molly Webb, Sarah Stout and Aaron Addison declare they have no competing interest.

Brian Ott: Grants and funds: Eli Lilly, Avid, Roche, TauRX, Merck, Univita, NIH/NIA; Honoria: NHTSA; Medscape; Consultant: Accera (DSMB).

David Carr: Grants and funds: Missouri Department of Transportation; Honorarium: Harvard Speaker; Consultant: Traffic Injury Research Foundation, Advanced Drivers Education Products and Training; Medscape.

John Morris: Grants/Funds: Healthy Aging and Senile Dementia, Antecedent Biomarkers for AD: The Adult Children Study, The Dominantly Inherited Alzheimer Network, and Alzheimer Disease Research Center. Honoraria: Cherkin Lecture, Chinese Society, 13th Eibsee Meeting (Keynote Speaker), Korean Dementia Association, and DZNE research center Magdeburg Symposium. Consultant: Lilly USA; ISIS Pharmaceuticals; Charles Dana Foundation. Royalties: Blackwell Medical Publishers; Taylor & Francis. Board Member: Board of Directors American Academy of Neurology (AAN).

Catherine Roe: Grants/Funds: NIH/NIA

Grant information

National Institute on Aging [R01AG043434, R01AG43434-03S1, P50-AG05681, P01-AG03991, P01-AG026276]; Fred Simmons and Olga Mohan, and the Charles and Joanne Knight Alzheimer's Research Initiative of the Washington University Knight Alzheimer's Disease Research Center.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Davis RL, Ohman JM: **Driving in Early-Stage Alzheimer's Disease: An Integrative Review of the Literature.** *Res Gerontol Nurs.* 2016; 1–15. [PubMed Abstract](#) | [Publisher Full Text](#)
- Duchek JM, Carr DB, Hunt L, *et al.*: **Longitudinal driving performance in early-stage dementia of the Alzheimer type.** *J Am Geriatr Soc.* 2003; 51(10): 1342–7. [PubMed Abstract](#) | [Publisher Full Text](#)
- Hunt LA, Murphy CF, Carr D, *et al.*: **Reliability of the Washington University Road Test. A performance-based assessment for drivers with dementia of the Alzheimer type.** *Arch Neurol.* 1997; 54(6): 707–12. [PubMed Abstract](#) | [Publisher Full Text](#)
- Odenheimer GL, Beaudet M, Jette AM, *et al.*: **Performance-based driving evaluation of the elderly driver: safety, reliability, and validity.** *J Gerontol.* 1994; 49(4): M153–9. [PubMed Abstract](#) | [Publisher Full Text](#)
- Ott BR, Davis JD, Papandonatos GD, *et al.*: **Assessment of driving-related skills prediction of unsafe driving in older adults in the office setting.** *J Am Geriatr Soc.* 2013; 61(7): 1164–9. [PubMed Abstract](#) | [Publisher Full Text](#)
- Babulal GM, Addison A, Ghoshal N, *et al.*: **Development and interval testing of a naturalistic driving methodology to evaluate driving behavior in clinical research [version 1; referees: 1 approved, 1 approved with reservations].** *F1000Research.* 2016; 5: 1716. [Publisher Full Text](#)
- Grengs J, Wang X, Kostyniuk L: **Using GPS data to understand driving behavior.** *J Urban Technol.* 2008; 15(2): 33–53. [Publisher Full Text](#)
- Blanchard RA, Myers AM, Porter MM: **Correspondence between self-reported and objective measures of driving exposure and patterns in older drivers.** *Accid Anal Prev.* 2010; 42(2): 523–9. [PubMed Abstract](#) | [Publisher Full Text](#)
- Wang X, Grengs J, Kostyniuk L: **Visualizing travel patterns with a GPS dataset: How commuting routes influence non-work travel behavior.** *J Urban Technol.* 2013; 20(3): 105–25. [Publisher Full Text](#)
- Kelly P, Krenn P, Titze S, *et al.*: **Quantifying the difference between self-reported and global positioning systems-measured journey durations: a systematic review.** *Transport Rev.* 2013; 33(4): 443–59. [Publisher Full Text](#)
- Molnar LJ, Eby DW: **The relationship between self-regulation and driving-related abilities in older drivers: an exploratory study.** *Traffic Inj Prev.* 2008; 9(4): 314–9. [PubMed Abstract](#) | [Publisher Full Text](#)
- Crizzle AM, Myers A, Vrkljan B, *et al.*: **Using in-vehicle devices to examine exposure and patterns in drivers with Parkinson's disease compared to an age-matched control group.** *Proc 6th int driving symp hum factor driver assess, training and vehicle design;* 2011. [Reference Source](#)
- National Highway Traffic Safety Administration: **Traffic Safety Facts 2012 data: Older population.** 2014. [Reference Source](#)
- National Highway Traffic Safety Administration: **Fatality Analysis Reporting System: Fatal Crash Trends 2012.** 2015. [Reference Source](#)
- Fan J, Han F, Liu H: **Challenges of Big Data Analysis.** *Natl Sci Rev.* 2014; 1(2): 293–314. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Eby DW, Silverstein NM, Molnar LJ, *et al.*: **Driving behaviors in early stage dementia: a study using in-vehicle technology.** *Accid Anal Prev.* 2012; 49: 330–7. [PubMed Abstract](#) | [Publisher Full Text](#)
- Marshall SC, Man-Son-Hing M, Bédard M, *et al.*: **Protocol for Candrive II/Ozcardrive, a multicentre prospective older driver cohort study.** *Accid Anal*

Prev. 2013; **61**: 245–52.

[PubMed Abstract](#) | [Publisher Full Text](#)

18. Skog I, Händel P: **Indirect instantaneous car-fuel consumption measurements.** *IEEE Transactions on Instrumentation and Measurement.* 2014; **63**(12): 3190–3198. [Publisher Full Text](#)
19. Wallace B, Goubran R, Knoefel F: **Cognitive change measurement through driving navigation ability sensing and analysis.** In *Medical Measurements and Applications Proceedings (MeMeA), 2013 IEEE International Symposium on.* 2013; 164–169. [Publisher Full Text](#)
20. Wallace B, Goubran R, Knoefel F, *et al.*: **Measuring variation in driving habits between drivers.** In *Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on.* 2014; 1–6. [Publisher Full Text](#)
21. Wallace B, Goubran R, Knoefel F, *et al.*: **Automation of the Validation, Anonymization, and Augmentation of Big Data from a Multi-year Driving Study.** *2015 IEEE International Congress on Big Data (BigData Congress).* 2015; 608–614. [Publisher Full Text](#)
22. Babulal G, Traub C, Webb M, *et al.*: **Dataset 1 in: Creating a driving profile for older adults using GPS devices and naturalistic driving methodology.** *F1000Research.* 2016. [Data Source](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 29 November 2016

doi:[10.5256/f1000research.10351.r17017](https://doi.org/10.5256/f1000research.10351.r17017)



Monique M. Williams

VITAS Healthcare, St. Louis, MO, USA

The authors describe their adaptation of a commercial off the shelf (COTS) in-vehicle device that captures GPS data. These data are analyzed using GIS techniques.

Data were collected from a total of five participants who were members of a longitudinal cohort study. The five participants were aged 65 years and older and were cognitively normal. They possessed a valid driver's license and drove at least once a week. In addition, all participants had provided Alzheimer's disease biomarker data (brain imaging or cerebral spinal fluid studies) in the past two years.

The authors used COTS Azuga G2 Tracking Device for data collection. The data included longitude, latitude, and vehicle speed. Data were collected every thirty seconds from the time that the ignition was turned on. In addition, aggressive driving behaviors such as speeding, hard braking, and sudden acceleration were logged. A driving profile was developed for each participant. The profiles included spatial, temporal, and behavioral components.

This pilot study demonstrates that the COTS device provides accurate data regarding daily driving behavior in older adults. The driving profiles for participants can be compared month-to-month or year-to-year and thus, provide the opportunity for observing changes in driving behavior.

With the aging of the population, the pilot study provides data that can be applied to the development of further studies to provide additional characterization of driving profiles of older adults.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 26 October 2016

doi:[10.5256/f1000research.10351.r17216](https://doi.org/10.5256/f1000research.10351.r17216)



Xueqin Qian

University of Minnesota, Department of Educational Psychology, Minneapolis, MN, USA

The authors adapted an in-vehicle device to measure driving behavior in individuals with Alzheimer's disease. Data from five participants were collected using the COTS Azuga G2 Tracking Device. Results from this study suggest that this technology was able to accurately identify locations and travel patterns of the participants.

Given the increasing number of individuals diagnosed with Alzheimer disease and the important role that driving plays in a person's life, this study has the potential to provide a measurement tool that can be used reliably to study driving involving seniors with Alzheimer disease. I have two recommendations for the authors to consider:

1. For the introduction, can authors provide a little background information on the driving behavior of senior citizens with Alzheimer? It may be helpful to orient readers who are unfamiliar with this topic.
2. As for results section, do you have data on under what situations, the alert behaviors happen the most frequently? If so, would it be possible to add it to the result section?

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Author Response 28 Nov 2016

Ganesh Babulal, Washington University School of Medicine, USA

Dr. Qian: Thank you for your prompt review and comments on this manuscript.

On your first point—since this is a methodological article, we limited the amount of information on driving performance and Alzheimer' disease in the background section. However, we have added a sentence in the introduction with a recent reference (2016) of a systematic review that summarizes the evidence on driving in early stage AD.

On your second point—we have the location (latitude and longitude), time and date of where/when the alert for adverse driving behaviors occurred. We are working to analyze these breadcrumbs, but given the high volume of data for each participant we do not have results available at this time that can be added into the result section of the current manuscript.

Competing Interests: No competing interests were disclosed.

Referee Report 21 October 2016

doi:[10.5256/f1000research.10351.r16603](https://doi.org/10.5256/f1000research.10351.r16603)



Frank-Dietrich Knoefel

Bruyère Research Institute, Ottawa, ON, Canada

With the current lack of an accepted standardized approach to assessing changing driving risk associated with aging, and with emerging technology, naturalistic driving is an important area of research. The authors of this paper have a strong track record in publishing in the field of aging, cognition and driving.

This paper provides the results of a pilot study of 5 participants using an off-the-shelf global positioning

data acquisition system connected to the OBDII port of the older driver's car. It describes how 5 months of data, collected by the company selling the device, was cleaned and processed to identify such features as spatial components (primary location, unique destinations and driving areas), and temporal components (time of day by season). They also used data obtained from the company on adverse driving behavior: speeding, hard breaking and hard acceleration. Using these features they were able to show differences between the drivers. It may be implied that changes in these parameters over time could be used in the future to help assess changing driving risk or help determine interventions to keep older adults driving longer, safely.

Major issues:

- While this paper is well-written and describes in some detail the work this group has done, it is not clear what the new contribution of this work is. While the Background suggests that the objective was to “describe methodological challenges associated with adapting a COTS in-vehicle device that captures and synthesizes GPS data for processing and analysis using GIS techniques” the abstract does not refer to this.
- Importantly, the design of this sensor system and analysis of data were not compared to other systems recently described in the literature, for instance by Marshall *et al.* (CanDrive project), Eby *et al.* (2012), and the SHRP2 study by Skog & Handel (2014). Specifically, work on destination identification has been published by Wallace *et al.* (2013). Our group has further addressed the issue of how to anonymize locations to protect individual driver identity (especially home address), a requirement of most research ethics committees (Wallace *et al.* 2015). Data cleaning has been discussed in papers by Porter and Wallace. The sensor information (breadcrumbs) logged every 30 seconds should be compared to other groups, for instance we sampled data every 1 to 5 seconds. The length of the pilot data is short compared to our published work using 1 year of driving data, from 7 years of available CanDrive data set. Finally, Wallace *et al.* (2014) have published work on driver identification that is not referenced
- Similarly, to be able to compare to other literature we need the definitions used for the alerts, e.g. how much over the speed limit was considered “speeding” (absolute or percent), and how this was determined (via GIS data?). Similarly we do not have definitions for hard accelerating and hard braking.

Minor issues:

- Background

Page 3: “recent studies are limited to capturing data... from weeks to 2 months.” CanDrive has collected driving data for 7 years from several hundred drivers – downloaded every 4 to 6 months.

- Methods

Unique destinations paragraph: for clarity: suggest “two or more distinct breadcrumbs occurring within the same radius during the same time period would be combined....”

I would move the sections describing the results of breadcrumbs (Fig 2) and the section how changing the buffer from 100 to 500 feet affects the grouping of breadcrumbs (Fig 3 and 4) into the Results section.

- Results:

The Abstract refers to 5 participants as a subset of the study participants, but we don't have a description of these drivers (age, gender, cognitive scores).

The first paragraph states: “A single trip could have hundreds of breadcrumbs that are aggregated and over time can provide specific information about driving patterns and behaviors.” Why were we not given the mean and range of breadcrumbs for trips, or time and distance of trips?

Figure 6 – legend: the lines for participants B and D look similar.

Page 7: Figure 6 is referred to twice – the second time should read Figure 7 (percentages).

References

1. Eby DW, Silverstein NM, Molnar LJ, LeBlanc D, Adler G: Driving behaviors in early stage dementia: a study using in-vehicle technology. *Accid Anal Prev*. 2012; **49**: 330-7 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Marshall SC, 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 MM, Rapoport M, Tuokko H, Vrkljan B, Woolnough A: Protocol for Candrive II/Ozdrive, a multicentre prospective older driver cohort study. *Accid Anal Prev*. 2013; **61**: 245-52 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Skog I., Händel P.: Indirect instantaneous car-fuel consumption measurements. *IEEE Transactions on Instrumentation and Measurement*. 2014; **63** (12): 3190-3198 [Reference Source](#)
4. Wallace B., Goubran R., Knoefel F.: Cognitive change measurement through driving navigation ability sensing and analysis. In *Medical Measurements and Applications Proceedings (MeMeA), 2013 IEEE International Symposium on*. 2013, May. 164-169 [Publisher Full Text](#)
5. Wallace B., Goubran R., Knoefel F., Marshall S, Porter M.: Measuring variation in driving habits between drivers. In *Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on*. 2014, June. 1-6 [Publisher Full Text](#)
6. Wallace B., Goubran R., Knoefel F., Marshall S., Porter S., Harlow M., Pauli A.: Automation of the Validation, Anonymization, and Augmentation of Big Data from a Multi-year Driving Study. *2015 IEEE International Congress on Big Data (BigData Congress)*. 2015, June. 608-614 [Publisher Full Text](#)

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 28 Nov 2016

Ganesh Babulal, Washington University School of Medicine, USA

Dr. Knoefel: We thank you for your thorough review and the additional references. We have made the following edits to the manuscript in consideration of your comments:

1. We have added in text into the abstract to clarify our objective for this paper.
2. We added information comparing the GPDAS device and data collection process to prior research studies that used naturalistic driving methodologies. The references have also been added as appropriate to the text.
3. We have added definitions for the alerts of speeding, hard braking and hard acceleration in the results section under the subsection, adverse driving behavior.
4. The sentence in the background section was modified to point out that some, not all studies have a short collection interval. We also referenced the CanDrive and its extensive data collection interval in the discussion section.
5. The clarification was made in the methods section to specific radius. We chose to retain the structure of the methods section to reflect a similar layout in the results section.

6. Demographics on the five participants were not included in this methodology article since the focus was on the GPDAS device, data collection and processing. Results examining differences among participants in driving behavior will be published later after more driving data is collected.
7. The mean and standard deviation for total number of trips and average miles per trip have been added for the group of participants.
8. We have also clarified the graph lines for the participants B and D in the text and corrected the reference to Figure 7 in the text.

Competing Interests: No competing interests were disclosed.
