CONTEXTUALIZING NATURALISTIC DRIVING DATA IN A RURAL STATE AMONG DRIVERS WITH AND WITHOUT OBSTRUCTIVE SLEEP APNEA

Jeffrey D. Dawson,[1] Elizabeth Bair[1], Nazan Aksan[2], Kelly Sewell[3], Jon Tippin,[2]
Matthew Rizzo,[4]
[1]Dept. of Biostatistics, Univ. of Iowa College of Public Health
Iowa City, Iowa, USA
[2]Dept. of Neurology, University of Iowa College of Medicine
Iowa City, Iowa, USA
[3]Dept. of Geographical and Sustainability Sciences,
Univ. of Iowa College of Liberal Arts and Sciences
Iowa City, Iowa, USA
[4]Dept. of Neurological Sciences, College of Medicine, University of Nebraska Medical Center
Omaha, Nebraska, USA
E-mail: jeffrey-dawson@uiowa.edu

Summary: In naturalistic studies, Global Positioning System (GPS) data and date/time stamps can link driver exposure to specific environments (e.g., road types, speed limits, night driving, etc.), providing valuable context for analyzing critical events, such as crashes, near crashes, and breaches of accelerometer limits. In previous work, we showed how to automate this contextualization, using GPS data obtained at 1 Hz and merging this with Geographic Information Systems (GIS) databases maintained by the Iowa Department of Transportation (DOT). Here we further demonstrate our methods by analyzing data from 80 drivers with obstructive sleep apnea (OSA) and 48 controls, and comparing the two groups with respect to several factors of interest. The majority of comparisons found no difference between groups, suggesting similar patterns of exposures to driving environments in OSA and control drivers. However, OSA drivers appeared to spend slightly more time on roads with annual traffic counts of 500-10,000 and less time driving on wider highways, during twilight, and on roads with 10,000-25,000 annual traffic counts.

INTRODUCTION

Naturalistic driving studies monitor participants as they operate their vehicles under everyday conditions. Study participants with impairments may, if they are aware of their impairments, modify their behavior by minimizing driving in high-risk situations. Thus, it is necessary to provide context when analyzing naturalistic driving data, to distinguish "exposure/strategy" (i.e., how often drivers put themselves in certain driving environments) from "safety/performance" (i.e., how the drivers actually operate their vehicle, given they are in such environments). Automating this process of linking Global Positioning System (GPS) data and date/time stamps can mitigate labor-intensive judgments based on video data.

Drivers with obstructive sleep apnea (OSA) tend to have higher risk of motor vehicle crashes than drivers without OSA (Tregear et al, 2009). The added risk likely depends on OSA severity,

compliance to treatment, and self-awareness of sleepiness (Engleman et al, 1997). We have been studying these issues in OSA drivers and controls. Our goals include comparing driving abilities and exposure strategies between groups, and evaluating how cognitive factors, measures of sleepiness, and treatment compliance relate to driving performance and strategies within the OSA group. This report summarizes and compares exposure strategies in OSA drivers and controls, by linking GPS data to Geographic Information System (GIS) maps and geocoded home addresses; using weather data from the national weather service; and using data/time stamps to determine natural lighting conditions.

METHOD

Subjects and study overview

The subjects in this report are 80 drivers with OSA and 48 comparison drivers, ages 30–60 years. All have at least 10 years of driving experience, use a single car as their primary vehicle (at least 90% of driving time), and drive at least 2 hours or 100 miles/week on average. All demographic characteristics were assessed at baseline. For this report, we focus on driving in the state of Iowa, where most of our drivers live, and where GIS data appears to be of greater availability than in surrounding states. We limit our main analyses to the first two weeks (out of ~3.5 months) of driving, as our OSA subjects typically began treatment for OSA after two weeks and we aimed to compare untreated OSA drivers to controls. The study was approved by the University of Iowa Institutional Review Board for Human Subjects Protection.

Driving monitoring and initial data preparation

Driving data were collected via electronic, video, and GPS outputs from a state-of-the-art instrumentation package installed in each participant's vehicle (McDonald et al, 2012) over a continuous 3.5-month period. For each trip, a 10-Hz file was created and uploaded with information pulled from the OBD2 port and from the accelerometers present in the installed device, and a 1-Hz file was created with GPS coordinate information. These two types of files were merged into one large comma-separated-value (CSV) text dataset per trip, and then concatenated into one dataset per subject. The amount of data collected, in terms of number of days, number of trips, rows of data, and data file size, varied from subject to subject.

As explained in our previous work (Dawson et al, 2015), each subject's 25-variable, 10-Hz data file was reduced to an eight-variable, 1-Hz file. These eight variable included GPS coordinates (latitude and longitude), a measure of GPS signal quality, a date/time stamp, and identification variables. Hence, the number of rows were reduced by 90%, and the number of columns by 68%, resulting in datasets mostly in the 10-40 Mb range.

Merging into GIS and other systems

In preparing to import these data into GIS software and databases, we first gave each data point a unique ID based on the driver ID and the observation count fields so it could be consistently referenced and tied back to the source data, and then eliminated data with null GPS information.

The remaining data were imported into a GIS, specifically ArcGIS 10.3 (ESRI 2013), and converted to a spatially explicit geodatabase feature. Next, clearly incorrect GPS information (e.g., coordinates outside of North America or unrelated to any drive path) were eliminated.

The GIS work was done in two programs: ArcMap (ArcGIS 10.1, ESRI 2013) and ArcGIS Pro (ESRI 2014). For consistency, efficiency, and accuracy, a workflow was created in ArcGIS ModelBuilder and saved as an automation tool. This tool was run on each subsequent dataset.

Road environment data curated by the state of Iowa are almost exclusively available as centerlines, or line vector data. Attributes that were determined to be of potential impact to a driver's decision-making process or abilities were parsed from statewide datasets (http://www.iowadot.gov/gis/downloads/zipped_files/GIMS_History/Statewide/). Due to a limit in the precision of our GPS, we buffered each road by 2 feet and used the Snap tool to overlay GPS points on the closest buffered road centerline. We then used the Spatial Join tool to connect the appropriate values of each underlying road layer to the GPS points.

The resulting datasets contained 70 new variables pertaining to road culture, including information on speed limit, 911-based street names, road surface type, etc. Hence, these datasets were roughly 10 times the size of the eight-variable datasets that were imported into the GIS software. The Iowa DOT of the state provided data dictionaries to aid in interpreting the values of the field. The processed data were then merged back into the original 25-variable files to be available for future formal analyses.

For this report, we analyzed six road environment factors that were of interest and appeared to be of reasonable completeness. Some potential variables of interest were eliminated based on suggestions from the DOT of purported reliable. Other variables were eliminated from consideration based on preliminary validation results based on video data. After eliminating such variables, we chose to examine the following:

      Speed limit (in MPH)
      Number of lanes (both directions)
      Average annual daily traffic
      Road system (Interstate, US Route, State Route, etc.)
      Area type (Residential, rural, etc.)
      Terrain (Flat, Rolling, Hilly)

Weather data came from an archive of the National Climactic Data Center's Automated Surface Observing System, (http://mesonet.agron.iastate.edu/ASOS/) maintained by Iowa State University. The weather data contained minute-to-minute observations generated from eighteen weather stations across the state of Iowa. We considered any of the following responses in the precipitation type variable as indicative of precipitation: R, R-, R+, S, S-, S+, or P.

Lighting condition data were composed of sunrise, sunset, civil twilight start, and civil twilight end times. Civil twilight is defined as when the center of the Sun is geometrically 6 degrees or less below the horizon. In the morning, twilight is the time between dawn and sunrise (approximately 30 minutes in Iowa); in the evening, twilight is the time between sunset and dusk (again, approximately 30 minutes). Using Iowa City, IA, as a reference point, these variables

were downloaded from the United States Naval Observatory website
(http://aa.usno.navy.mil/data/docs/RS_OneDay.php) and then used to categorize twilight, daylight, and nighttime categories of the lighting condition factor.

Home addresses were geocoded with the U.S. Census Bureau Geocoder
(https://geocoding.geo.census.gov/) using the public address ranges current benchmark. Straight line distance between the home address and location while driving was subsequently calculated using the haversine formula.

Statistical analyses

Mean and standard deviation or count and percentage were calculated for each of the demographic characteristics as appropriate. OSA drivers and controls were compared using two-sample t-tests, the Wilcoxon rank-sum test (a.k.a., Mann-Whitney U tests), or Fisher's exact test.

For each driver, we calculated the number of drives, the number of days of observation (up to 14), the total drive time (in hours), and the mean length per drive (in minutes). For each category of each factor of interest (six road environment variables, precipitation, time of day category, and distance from home), we calculated the proportion of time that each subject drove in those conditions. For example, if a subject drove a total of 20 hours in the first 14 days, and 3 of those hours were on an interstate, then the "Interstate" category of the "Road System" factor would have a value of 0.15 for that driver.

Once the data were reduced to one summary per driver for each variable category, we calculated means, standard deviations, and medians as descriptive statistics. We then compared OSA drivers to controls using two-sample t-tests or Wilcoxon rank sum tests, depending on whether the distribution of proportions appeared to be reasonable normal.

RESULTS

As detailed in Table 1, the OSA driver group did not differ from the control driver group in terms of age, gender, or Hispanic ethnicity, but were significantly different in terms of race and education. More OSA drivers identified as African American/Black than controls, while more controls identified as Other. Additionally, OSA drivers were overall less educated than controls.

In the entire course of the study, our 80 OSA drivers and 48 controls had a total of 45,019 drives and 9,701.9 hours of driving in the state of Iowa. In the first 14 days of the study (our focus), the OSA drivers had 4,277 total drives and the controls had 2,751 total drives. As can be seen in Table 2, the distribution of number of drives per subject, duration of subject observation, total drive time, and mean length of drive were similar between the two groups of drivers.

The comparisons of our nine factors of interest (35 total variables) can be seen in Table 3. The majority of comparisons showed similar distributions between the two groups. Five of the nine factors had no significant or near-significant differences at all, namely, speed limit, road system, terrain, precipitation, and distance from home.

Table 1. Driver demographic characteristics

| Variable | Mean (SD) or N (%) | | p-value |
	OSA (n = 80)	Control (n = 48)	
Age	46.0 (8.0)	44.2 (8.6)	0.2189[a]
Gender			0.6671[b]
Male	53 (66.25)	30 (62.50)	
Female	27 (33.75)	18 (37.50)	
Race			0.0036[c]
White	61 (76.25)	36 (75.0)	
African American/Black	10 (12.50)	0 (0.00)	
Asian	6 (7.50)	3 (6.25)	
American Indian/Alaska Native	1 (1.25)	2 (4.17)	
Other	2 (2.50)	7 (14.58)	
Ethnicity			0.3675[c]
Hispanic	6 (7.59)	6 (12.5)	
Education			0.0116[b]
High School or less	14 (17.50)	2 (4.17)	
Vocational/Professional training or Associate's Degree	33 (41.25)	12 (25.00)	
Bachelor's Degree and possibly some professional school	20 (25.00)	15 (39.58)	
Graduate Degree	13 (16.25)	19 (31.25)	

[a] t-test
[b] Wilcoxon rank sum test
[c] Fisher's exact test

Table 2. Characteristics of 14 days of driving in Iowa

| Variable | Mean (SD), Median | | p-value |
	OSA (n = 80)	Control (n = 48)	
Number of drives per subject	53.5 (35.5), 49	57.3 (41.1), 55	0.6436[a]
Duration of subject observation (days)	10.9 (3.9), 13.0	12.0 (2.7), 13.0	0.36859[a]
Total drive time observed (in hours) by subject	11.0 (6.9), 11.2	11.9 (7.8), 12.6	0.4954[a]
Mean length (in minutes) of subject drives	13.3 (6.4), 11.9	12.7 (5.5), 10.8	0.5017[a]

[a] Wilcoxon rank sum test

There were four variables where there appeared to be differences (p<0.05). OSA drivers spent more time proportionally on roads with annual traffic counts of 500-10,000, and less time proportionally on roads that had 6 or more lanes, on roads with annual traffic counts of 10,000-25,000, and driving during twilight. There were also two variables of near-statistical significance (0.05<p<0.10), suggesting that OSA drivers may have driven less in fringe business districts or rural areas.

Table 3. Proportion of drive time in Iowa occurring during particular road environment, precipitation, and light conditions

Variable	Category	Mean (SD), Median		p-value
		OSA (n = 80)	Control (n = 48)	
Speed limit	10-25 mph	0.281 (0.178), 0.243	0.294 (0.135), 0.261	0.2232[b]
	30-35 mph	0.188 (0.121), 0.163	0.212 (0.112), 0.200	0.1314[b]
	40-50 mph	0.082 (0.043), 0.075	0.085 (0.048), 0.074	0.8151[b]
	55-60 mph	0.182 (0.161), 0.132	0.156 (0.128), 0.127	0.6956[b]
	65-70 mph	0.180 (0.166), 0.138	0.159 (0.152), 0.108	0.5899[b]
Number of lanes	1-3 lanes	0.490 (0.177), 0.493	0.477 (0.166), 0.481	0.6912[a]
	4-5 lanes	0.365 (0.160), 0.359	0.355 (0.140), 0.350	0.7034[a]
	6 or more lanes	0.055 (0.036), 0.050	0.073 (0.048), 0.064	0.0283[b]
Average annual daily traffic	1-500	0.079 (0.068), 0.060	0.084 (0.096), 0.065	0.8574[b]
	500-10,000	0.429 (0.170), 0.430	0.352 (0.153), 0.319	0.0111[a]
	10,000-25,000	0.229 (0.105), 0.228	0.271 (0.105), 0.277	0.0274[a]
	>25,000	0.173 (0.149), 0.129	0.188 (0.155), 0.149	0.4954[b]
Road system	Interstate	0.153 (0.172), 0.089	0.156 (0.178), 0.077	0.9784[b]
	U.S. Route	0.180 (0.161), 0.120	0.165 (0.134), 0.125	0.8904[b]
	State Route	0.084 (0.111), 0.056	0.090 (0.088), 0.072	0.4267[b]
	Farm to Market Route	0.192 (0.145), 0.171	0.188 (0.110), 0.176	0.7102[b]
	Local Road	0.391 (0.198), 0.351	0.402 (0.147), 0.414	0.4470[b]
Area type	Central Business District	0.017 (0.020), 0.010	0.015 (0.015), 0.012	0.8767[b]
	Fringe Business District	0.048 (0.038), 0.039	0.067 (0.066), 0.045	0.0598[b]
	Outlying Business District	0.141 (0.084), 0.128	0.137 (0.087), 0.127	0.7923[b]
	Residential Area	0.233 (0.148), 0.211	0.242 (0.116), 0.226	0.7154[a]
	Rural area	0.236 (0.169), 0.196	0.267 (0.125), 0.261	0.0522[b]
Terrain	Flat	0.133 (0.131), 0.098	0.094 (0.094), 0.065	0.1475[b]
	Rolling	0.132 (0.135), 0.091	0.115 (0.122), 0.059	0.4814[b]
	Hilly	0.029 (0.047), 0.012	0.021 (0.029), 0.011	0.8317[b]
Precipitation	No Precipitation	0.945 (0.113), 0.982	0.961 (0.054), 0.993	0.4502[b]
	Precipitation	0.055 (0.113), 0.018	0.039 (0.054), 0.007	0.4502[b]
Lighting condition	Twilight	0.051 (0.084), 0.027	0.074 (0.081), 0.041	0.0325[b]
	Daytime	0.785 (0.181), 0.825	0.734 (0.207), 0.728	0.2032[b]
	Nighttime	0.164 (0.157), 0.131	0.191 (0.180), 0.166	0.5530[b]
Distance from home§	≤1 mile	0.211 (0.156), 0.178	0.202 (0.148), 0.189	0.7273[b]
	1-2 miles	0.128 (0.107), 0.112	0.152 (0.099), 0.140	0.1386[b]
	2-5 miles	0.210 (0.147), 0.179	0.230 (0.142), 0.227	0.3369[b]
	5-10 miles	0.133 (0.124), 0.104	0.131 (0.110), 0.113	0.7826[b]
	>10 miles	0.317 (0.258), 0.293	0.284 (0.261), 0.188	0.5166[b]

Note: Proportions may not add to 1.0, due to missing data and/or rounding.

[a] t-test

[b] Wilcoxon rank sum test

§ Due to inadequate address information, OSA: n = 76 and Control: n = 47

DISCUSSION

Overall, the patterns of driving in OSA subjects were similar to those of drivers without OSA. Among the 35 comparisons involving our 9 factors of interest, we found four significant differences, somewhat higher than the 1.75 expected number of false positive findings (Type I errors) if all of null hypotheses were true (i.e., 35 x 5% = 1.75). Hence, there appear to be actual

differences between the two groups, albeit few. The total number of drives and the length of drives were similar in the two groups. Thus, untreated OSA drivers in this study appear to have similar exposures as controls and may not be making large strategic adjustments in their driving.

Our study has a number of limitations. First, we were limited by the accuracy and precision of the GIS and weather databases, as well as our GPS devices. Second, we only used one location for information on sunrise, sunset, and twilight, which would lead to misclassifications when the vehicles are in locations far from this location. Third, there may be other factors of interest, such as road curvature, which have not yet been obtained or analyzed. Finally, we did not incorporate multiple residential addresses or work addresses in our analyses.

Future analyses of these data include examining trends of weekend vs. weekday driving, driving during rush hours, and seasonal effects. We will also test if CPAP treatment of OSA drivers affects the factors we examined in this study. We are in the process of validating key factors of interest using video clips available in this study. Ultimately, we will adjust for the contextualization variables when we measure how driving safety and performance are affected by disease status, treatment, and cognitive factors. Thus, we will be able to measure exposure in meaningful new ways, to better understand person-level factors, exposure, and driving behavior.

We are applying these methods to an ongoing study of healthy elderly drivers. This new study has drivers who drive more in the western part of Iowa as well as in other states. Of note, the amount and quality of data pertaining to road characteristics varies greatly across states. This issue results in great challenges when analyzing multi-state naturalistic data. For certain factors, such as lighting condition, road type, and speed limit, it may be possible to use high-definition video and computer vision algorithms to classify road segments in a more automated manner.

ACKNOWLEDGMENTS

REFERENCES

Dawson, J.D., Yu, L., Sewell, K., Skibbe, A., Aksan, N., Tippin, J., & Rizzo, M. (2015). Linking GPS data to databases to assess driving patterns in drivers with obstructive sleep apnea. *Proceedings of Driving Assessment 2015: The Eighth International Driving Symposium on Human Factors in Driving Assessment, Training and Vehicle Design*, 147-153.

Engleman, H.M., Hirst, W.S.J., & Douglas, N.J. (1997). Under reporting of sleepiness and driving impairment in patients with sleep apnea/hypopnea syndrome. *Journal of Sleep Research*, 6, 272-275.

McDonald, A.D., Lee, J.D., Aksan, N.S., Rizzo, M., Dawson, J.D., & Tippin, J. (2012). Making Naturalistic Driving Data SAX-y. *Proceedings of the VTTI Third International Symposium on Naturalistic Driving Research*, Blacksburg, Virginia.

Tregear S, Reston J, Schoelles K, & Phillips B. (2009). Obstructive sleep apnea and risk of motor vehicle crash: systematic review and meta-analysis. *J Clin Sleep Med*, *5*(6), 573-581.