

## Understanding Health and Well-Being from Naturalistic Driving Behavior

Han Zhang<sup>1</sup>, Manhua Wang<sup>2</sup>, David W. Eby<sup>1</sup>, Renée St. Louis<sup>1</sup>

<sup>1</sup> University of Michigan Transportation Research Institute

<sup>2</sup> Department of Industrial and Operations Engineering, University of Michigan

---

### Author Note:

Correspondence to this manuscript should be addressed to Han Zhang (hanzh@umich.edu).

Data and code associated with this manuscript can be found at <https://github.com/HanZhang-psych/Understanding-Health-and-Well-Being-from-Naturalistic-Driving-Behavior>

Author roles: Han Zhang: Conceptualization, formal analysis, visualization, Writing – original draft. Manhua Wang: Writing – review & editing. David W. Eby: Funding acquisition, Resources, Project administration, Writing – review & editing. Renée St. Louis: Funding acquisition, Resources, Project administration, Writing – review & editing.

The authors have no known conflicts of interest to disclose.

## **Abstract**

The way humans interact with their environment leaves traces that reflect their health and well-being. Advances in sensing technologies now make it possible to capture these traces passively and unobtrusively. Here, we test the idea of using everyday driving patterns to predict health and well-being. We analyzed everyday driving patterns from 2,658 older adults who also completed self-reported measures of general well-being and health across cognitive, physical, social, and mental domains. Driving behaviors predicted both overall well-being and specific health domains beyond demographic variables. Physical and social health were most strongly associated with driving variables. Furthermore, we identified several driving signatures of health that were highly specific in their predictions. Finally, including driving variables improved out-of-sample predictive performance relative to a demographic-only model. Overall, these findings position driving behavior as a potential new source of personal sensing that can be leveraged to understand and promote health and well-being.

Recent advances in sensing technologies have transformed how researchers study human behavior, making it possible to unobtrusively collect fine-grained data in naturalistic settings<sup>1,2</sup>. Meanwhile, growing debates in psychology and related fields have highlighted the need for more ecologically valid, *in situ* measures of behavior to understand mental processes<sup>3,4</sup>. These trends have converged with the emerging need in healthcare to develop ongoing, individualized symptom monitoring systems that extend beyond periodic clinical visits<sup>5,6</sup>. Together, these forces have given rise to the growing field of *personal sensing*, which harnesses passively collected data from everyday life to better understand and promote health and well-being<sup>1,5,7</sup>.

To date, personal sensing research has predominantly focused on personal digital devices such as smartphones and wearables<sup>1,5,7</sup>. The widespread accessibility of these devices and their various built-in sensors make them a good option for personal sensing. Data such as GPS location, phone interactions, heart rate, and voice recordings have been used to monitor health behaviors<sup>8,9</sup>, characterize social contexts<sup>10,11</sup>, and identify risks for mental health conditions<sup>2,12,13</sup>.

Yet, personal digital devices are not the only source of sensor data that can shed light on health and well-being. In the United States, driving is almost universal: In 2023, an estimated 258.2 million people in the US, or 95.3% of those aged 16 or older, drove a vehicle<sup>14</sup>. People spent an average of 60.7 minutes per day behind the wheel for various purposes, such as commuting to work, running errands, attending social activities, and medical appointments. The need for driving cuts across demographic groups such as age, race, sex, geographic locations, and education levels. At the same time, driving is a complex task that requires significant physical, perceptual, and mental abilities for maneuvers, control, and planning<sup>15,16</sup>. The prevalence of driving, its deep integration into many aspects of daily life (at least in the US), and the range of abilities it demands suggest that driving behaviors, too, may carry information about individuals' health and well-being.

Psychological and health variables in transportation research have been examined primarily through the lens of how these factors influence driving performance and safety risks<sup>17–19</sup>, often based on self-reported driving behaviors<sup>20</sup>. Over the past decade, there has been a shift toward large-scale naturalistic driving datasets that record months to years of everyday driving<sup>21–23</sup>. Installed sensors can capture continuous, real-world driving behaviors that extend beyond safety-critical events (e.g., accidents and traffic violations) to include broader patterns such as in-vehicle driver behaviors, vehicle kinematics, travelled locations, and many more. Building on this, a small but growing body of research has begun to take the reverse perspective and ask the question: Can we leverage a person's driving pattern to understand that person's health and well-being<sup>24–26</sup>? These early studies, although limited to single health outcomes, show promising results and highlight the need for broader investigations to determine whether specific driving behaviors are unique markers of certain health conditions or general indicators of overall well-being.

Here, we studied one of the largest naturalistic driving samples to date, which includes everyday driving patterns of 2,658 older drivers in the US from 2015 to 2019<sup>21</sup>. Focusing on older adults is especially important given the rapidly aging driving population and the growing need to monitor health outcomes to support healthy aging<sup>27</sup>. We examined how everyday driving behaviors relate to a comprehensive set of health and well-being domains, including cognitive, physical, social, and mental health (see Table 1). We also identified distinct driving markers uniquely linked to specific domains by modeling multiple health measures simultaneously and comparing the strength of relationships. Through this work, we seek to establish the connection between driving behaviors and multiple aspects of health and well-being.

**Table 1**

*A summary of health and well-being measures examined in this study.*

Measure	Definition
<b>GENERAL</b>	
Life Satisfaction	Overall satisfaction with life.
<b>COGNITIVE</b>	
Cognitive Decline	Perception of one's cognitive issues in areas such as concentration, memory, and mental acuity.
<b>PHYSICAL</b>	
Physical Decline	Perception of one's physical issues such as having difficulty doing chores, walking up and down stairs, and running errands.
Fatigue	Feelings of tiredness and exhaustion.
<b>SOCIAL</b>	
Role Constraints	Perceived difficulty in performing one's usual roles and activities such as working and hanging out with family and friends.
Social Isolation	Perceptions of being isolated, excluded, disconnected from, or unknown by others.
Informational Support	Perceived availability of helpful information or advice when needed.
Emotional Support	Feelings of being cared for and valued as a person; having confident relationships.
Instrumental Support	Perceived availability of material, task, or medical assistance when needed.
<b>MENTAL</b>	
Depression	Feeling hopeless, worthless, helpless, or sad.
Anxiety	Feeling anxious, fearful, or overwhelmed by worry.
Anger	Feeling angry, annoyed, grouchy, or ready to explode.

## Results

We analyzed data from 2,658 participants in the AAA Longitudinal Research on Aging Drivers (AAA LongROAD) study<sup>21</sup>. Participants aged 65 to 79 who were active drivers were recruited from five study sites across the United States. Sample characteristics are presented in Table 2. At the baseline visit, a device was installed in each participant's primary vehicle to record driving data as participants drove their vehicles for daily travel. The raw driving data were cleaned to compute 24 monthly driving measures. See Table S1

in the supplemental material for definitions of the driving variables, Figure S1 for distributions, and Figure S2 for a correlation matrix of driving variables. A median of 29 full months of driving data was available for each participant (range: 3–38 months). Participants also completed annual assessments at baseline and up to two follow-up sessions, during which they completed questionnaires assessing general well-being (overall life satisfaction) and specific health domains, including cognitive, physical, social, and mental health. Because the present study focused on driving–health associations at the individual level, participant-level averages were calculated by aggregating driving data across months and health data across sessions.

**Table 2**

*Demographic information of the study participants.*

	Number (%) of Participants (N=2658)
<b>GEOGRAPHIC LOCATION</b>	
Cooperstown, New York	536 (20.2%)
Baltimore, Maryland	534 (20.1%)
Denver, Colorado	506 (19.0%)
La Jolla, California	523 (19.7%)
Ann Arbor, Michigan	559 (21.0%)
<b>AGE</b>	
Mean (SD)	71.05 (4.07)
<b>SEX</b>	
Male	1261 (47.4%)
Female	1397 (52.6%)
<b>RACE</b>	
White, Non-Hispanic	2283 (85.9%)
Black, Non-Hispanic	192 (7.2%)
American Indian	17 (0.6%)
Asian	57 (2.1%)
Alaska Native, Native Hawaiian, Pacific Islander	9 (0.3%)
Other, Non-Hispanic	28 (1.1%)
Hispanic	72 (2.7%)
<b>EDUCATION</b>	
1st-8th grade	8 (0.3%)

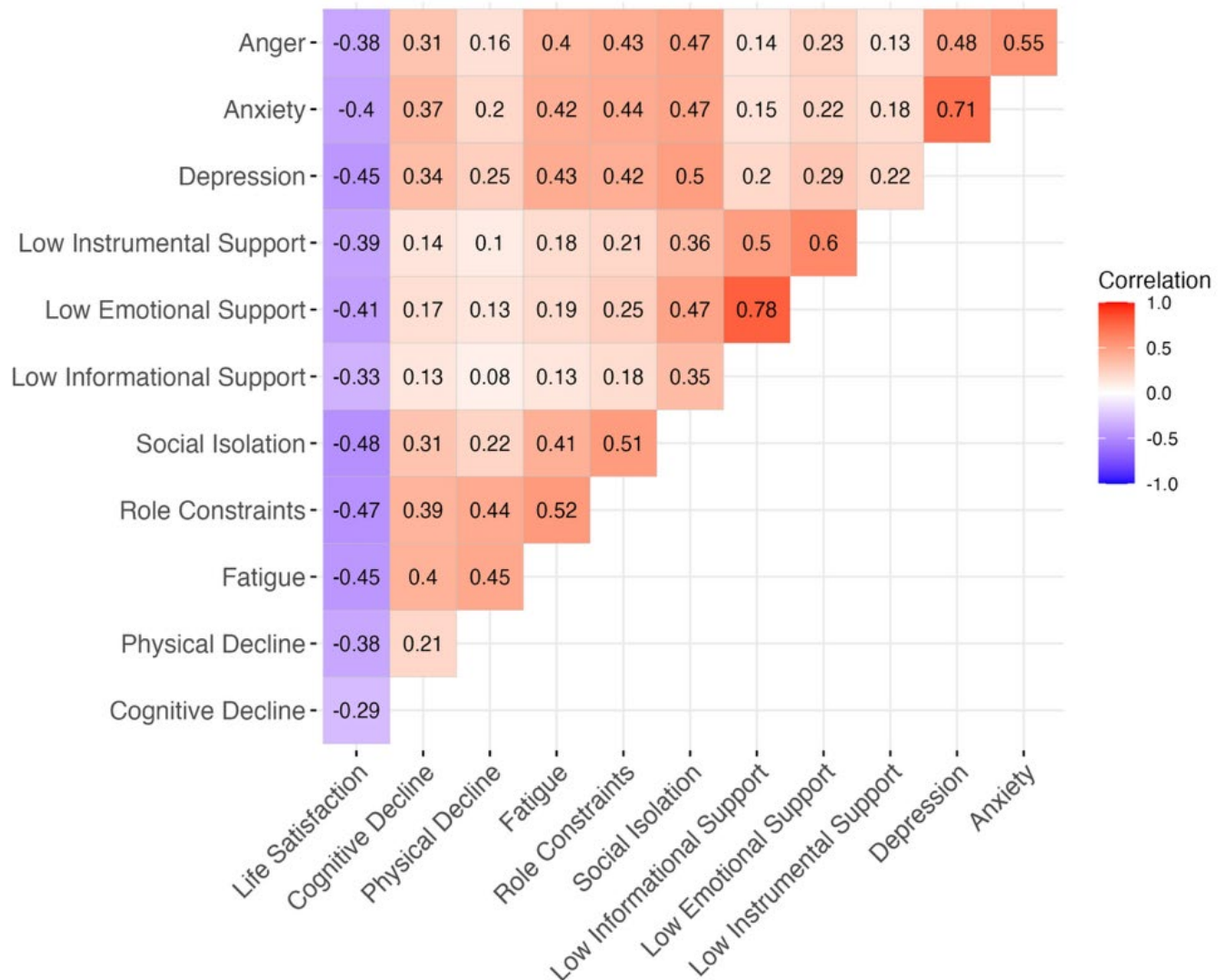
	Number (%) of Participants (N=2658)
9th-12th grade (no diploma)	39 (1.5%)
High school graduate (high school diploma or equivalent)	238 (9.0%)
Vocational, technical, business, or trade school (beyond high school level)	72 (2.7%)
Some college but no degree	392 (14.7%)
Associate degree	181 (6.8%)
Bachelor degree	622 (23.4%)
Master, professional, or doctoral degree	1106 (41.6%)
<b>HOUSEHOLD INCOME</b>	
Less than \$20,000	111 (4.2%)
\$20,000 to \$49,999	587 (22.1%)
\$50,000 to \$79,999	671 (25.2%)
\$80,000 to \$99,999	394 (14.8%)
\$100,000 or more	895 (33.7%)
<b>CURRENTLY WORKING</b>	
No	1839 (69.2%)
Yes	819 (30.8%)
<b>MARRIAGE</b>	
Married	1694 (63.7%)
Living with a partner	92 (3.5%)
Separated	33 (1.2%)
Divorced	388 (14.6%)
Widowed	337 (12.7%)
Never married	114 (4.3%)

### Driving Behaviors Predicted Overall Life Satisfaction

Overall life satisfaction showed consistent negative correlations with health measures (see Figure 1), supporting its use as an indicator of general well-being. We regressed life satisfaction on each driving variable, controlling for geographic location, age, sex, race, education, income, work status, and marital status. These multiple regression models, therefore, assess whether there is any utility in including a driving variable to understand general well-being beyond what can be inferred from a person's basic demographic characteristics.

**Figure 1**

*A correlation matrix of health and well-being measures.*



We found that life satisfaction was positively predicted by travel speed (standardized  $\beta = 0.114$ , 95% CI [0.072, 0.157],  $t(2636) = 5.28$ , Bonferroni-adjusted  $p < .001$ ), trips during morning rush hours ( $\beta = 0.073$ , 95% CI [0.036, 0.109],  $t(2636) = 3.94$ ,  $p = .02$ ), and the percentage of trips during morning rush hours ( $\beta = 0.104$ , 95% CI [0.068, 0.140],  $t(2636) = 5.61$ ,  $p < .001$ ). In contrast, life satisfaction was negatively predicted by the percentage of trips during evening rush hours ( $\beta = -0.100$ , 95% CI [-0.137, -0.064],  $t(2636) = -5.37$ ,  $p < .001$ ) and the number of hard-braking events per 1,000 miles ( $\beta = -0.080$ , 95% CI [-0.116, -0.043],  $t(2636) = -4.25$ ,  $p = .006$ ).

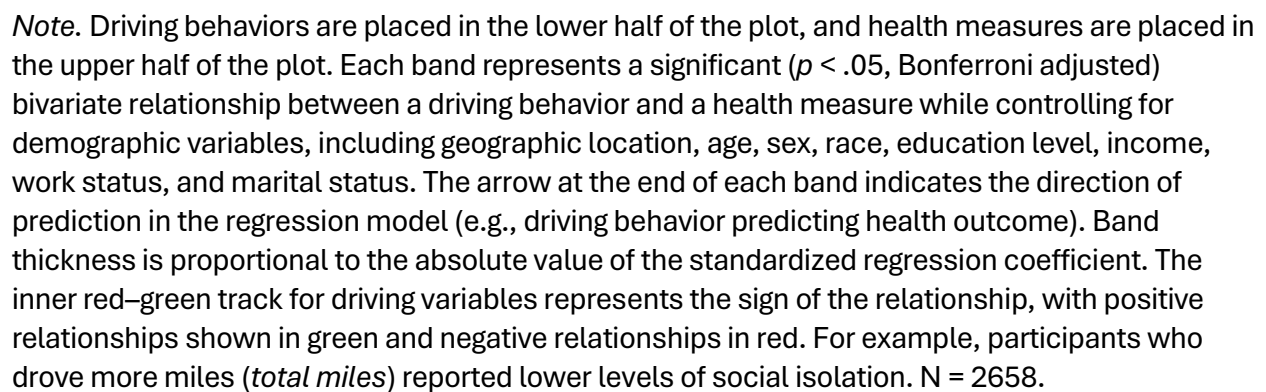
## **Widespread Associations Between Driving Behaviors and Health Domains**

Next, we examined how specific health domains are related to driving behaviors. Each health measure was regressed on a single driving variable, controlling for geographic location, age, sex, race, education level, income, work status, and marital status. Significant bivariate associations are visualized in Figure 2, and a numerical summary is provided in Table 3. All health measures except for informational support, emotional support, and anger were significantly associated with at least one driving behavior. Physical decline had the highest number of significant predictors (10), followed by social isolation (4), role constraints (3), depression (3), anxiety (2), fatigue (2), and cognitive decline (1). Of the 24 driving behaviors examined, 15 were significantly associated with at least one health measure.

Several driving variables were uniquely associated with a single health measure. For example, participants who had a higher number of days driving reported lower levels of physical decline, whereas those who spent more minutes per trip reported higher levels of physical decline. In contrast, some driving variables were associated with multiple health measures. For example, participants who had a lower travel speed reported greater physical decline, anxiety, role constraints, depression, and fatigue.



*Significant bivariate relationships between driving behavior and health measures.*



**Table 3**

*A numeric summary of significant bivariate relationships between driving behavior and health measures.*

Measures	Standardized $\beta$	95% CI	$t$ ( $df = 2636$ )	$p$ value
<b>COGNITIVE DECLINE</b>				
Miles Per Chain	-0.071	[-0.108, -0.034]	-3.79	0.045
<b>PHYSICAL DECLINE</b>				
Travel Speed (MPH)	-0.137	[-0.181, -0.093]	-6.10	< .001
Number of Days Driving	-0.121	[-0.158, -0.084]	-6.37	< .001
Number of Left Turns	-0.078	[-0.116, -0.039]	-3.96	0.022
Minutes Per Chain	0.073	[0.035, 0.111]	3.79	0.044
Minutes Per Trip	0.102	[0.063, 0.140]	5.20	< .001
Percentage of Trips AM Rush Hour	-0.100	[-0.138, -0.062]	-5.20	< .001
Total Trip Chains	-0.091	[-0.129, -0.052]	-4.63	0.001
Total Trips	-0.121	[-0.158, -0.083]	-6.33	< .001
Number of Trips AM Rush Hour	-0.117	[-0.154, -0.079]	-6.11	< .001
Number of Trips < 15 Miles	-0.115	[-0.152, -0.078]	-6.03	< .001
<b>FATIGUE</b>				
Travel Speed (MPH)	-0.100	[-0.145, -0.055]	-4.34	0.004
Percentage of Trips PM Rush Hour	0.079	[0.040, 0.118]	3.99	0.019
<b>ROLE CONSTRAINTS</b>				
Travel Speed (MPH)	-0.118	[-0.163, -0.073]	-5.14	< .001
Percentage of Trips AM Rush Hour	-0.078	[-0.117, -0.039]	-3.96	0.022
Percentage of Trips PM Rush Hour	0.108	[0.069, 0.147]	5.45	< .001
<b>SOCIAL ISOLATION</b>				
Number of Left Turns	-0.078	[-0.117, -0.038]	-3.87	0.031
Total Miles	-0.092	[-0.132, -0.052]	-4.49	0.002
Total Trip Minutes	-0.092	[-0.131, -0.052]	-4.57	0.001
Total Trips	-0.078	[-0.117, -0.04]	-4.02	0.017
<b>DEPRESSION</b>				
Travel Speed (MPH)	-0.112	[-0.157, -0.067]	-4.89	< .001
Total Miles	-0.078	[-0.118, -0.038]	-3.81	0.041
Miles Per Chain	-0.077	[-0.116, -0.037]	-3.82	0.040
<b>ANXIETY</b>				
Travel Speed (MPH)	-0.120	[-0.165, -0.075]	-5.21	< .001
Miles Per Trip	-0.080	[-0.120, -0.039]	-3.87	0.032

Notes.  $p$  values were adjusted using the Bonferroni method.

## Driving Behaviors Varied in the Specificity of Predictions

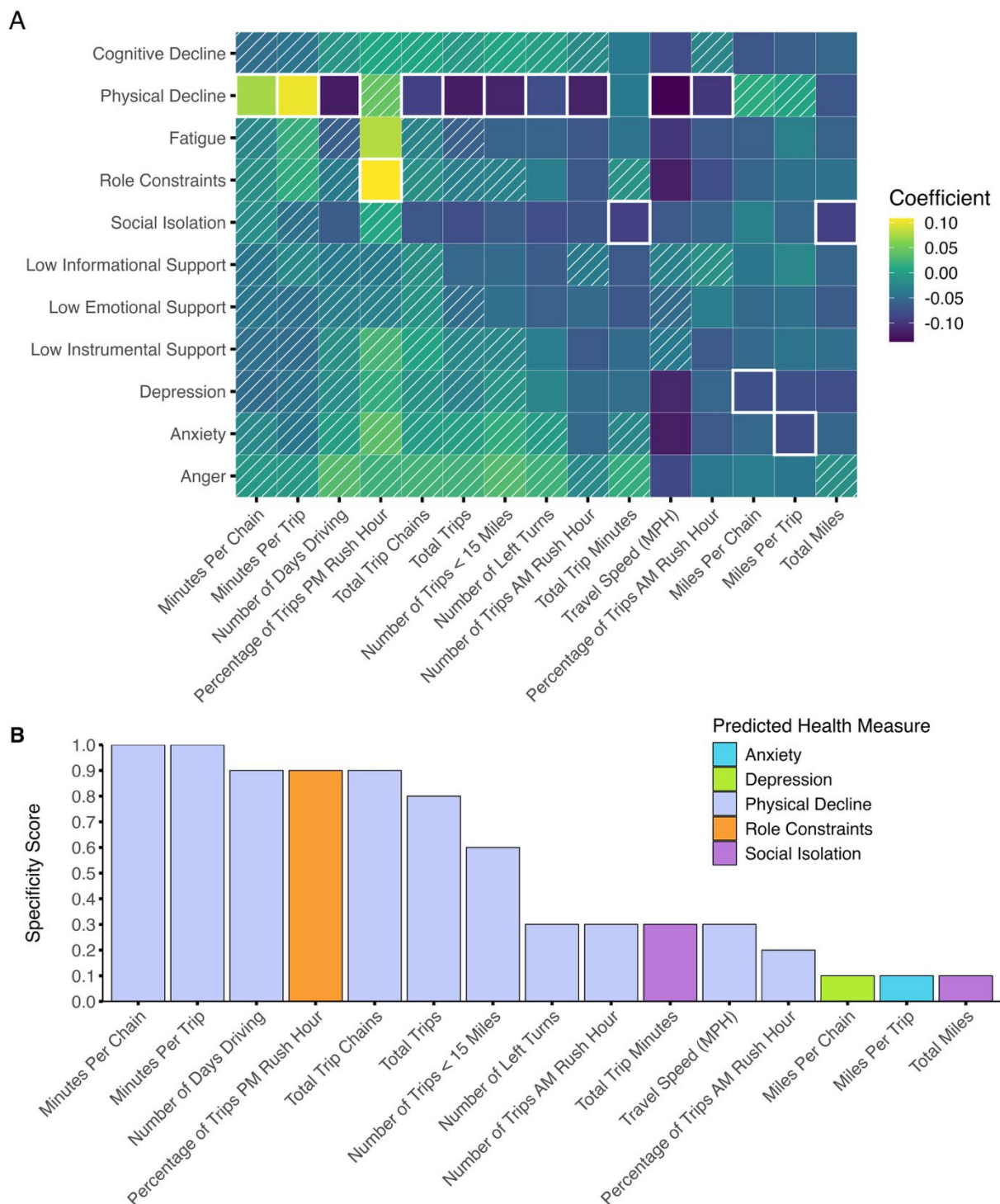
To further identify behavioral markers of health, we constructed multivariate path models in which each driving variable's associations with all health measures were directly compared. For each variable, we selected a candidate relationship to be the reference to compare against all other associations. For variables with a single significant association (e.g., *minutes per trip*), that association was selected as the candidate relationship. For variables significantly associated with multiple health measures, the strongest association was chosen as the candidate relationship. For example, for *travel speed*, its association with physical decline was the candidate relationship, given that it was numerically the strongest association (see Figure 2).

Figure 3A summarizes these comparisons. A complete summary of the results can be found in Table S2. For each driving variable, the candidate relationship is highlighted with a white rectangle. A crossed-out tile indicates that the candidate relationship dominated (i.e., was significantly stronger than) the association represented by that tile, as determined by the path model analysis. A specificity score for each driving variable was calculated as the percentage of relationships it significantly dominated (see Figure 3B). A higher specificity score indicates that the relationship was not only significant by itself but was also significantly stronger than the relationships of the same driving variable with other health measures.

Driving variables varied in their specificity of predictions. The most specific predictors were *minutes per chain* and *minutes per trip*, both of which were more strongly associated with physical decline than with other health measures. The number of days driving and the total number of trip chains are also highly specific predictors of physical decline. Both variables were more strongly associated with physical decline than other health variables, except for social isolation. Another highly specific predictor was the percentage of trips completed during evening rush hour, which was more strongly associated with role constraints than other health measures except fatigue. By contrast, some variables, such as travel speed, were associated with multiple health outcomes and were not specific to a single aspect of health.

To further investigate the high specificity of *minutes per chain* and *minutes per trip* in predicting physical decline, we conducted a supplemental analysis in which these variables were used to predict physical decline while controlling for *miles per chain* or *miles per trip*. The idea was that, if individuals with greater physical decline require more time to complete each trip or chain, the association between time-based metrics and physical decline should become stronger once the distance traveled is held constant. This is exactly what we observed (see Table S3 in the Supplement): holding miles constant, individuals who required more minutes per chain or trip reported higher levels of physical decline.

**Figure 2**  
*Dominance matrix (panel A) and specificity score (panel B) of driving behaviors.*



*Note.* Panel A presents a matrix of driving behaviors and health measures, with tile color indicating the strength of each bivariate relationship. The tile outlined with a rectangle marks the candidate relationship, which serves as the reference for comparisons within the same column. A crossed-out tile indicates that the candidate relationship was significantly stronger than the relationship represented by that tile. All tests were controlled for the same demographic variables as those in the bivariate analysis. Panel B displays the specificity score for each driving variable shown in Panel A, calculated as the percentage of relationships dominated by the candidate relationship. A higher specificity score indicates that the relationship is not only significant by itself but also significantly stronger than the relationships of the same driving variable with other health measures.  $N = 2658$ .

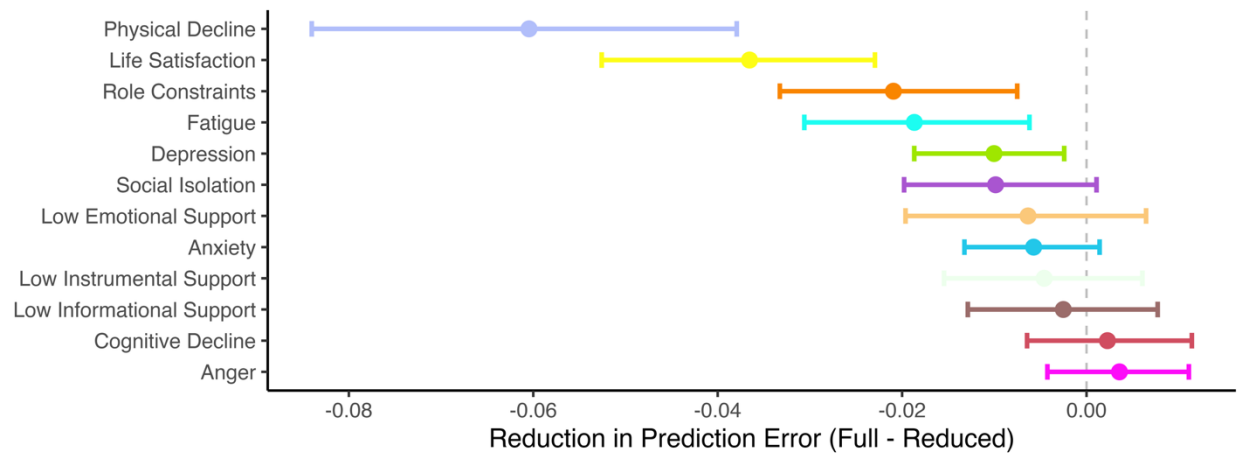
### **Driving Variables Improved Out-of-Sample Prediction of Health and Well-Being**

Finally, we tested whether driving variables, as a whole, improved out-of-sample predictions of health and well-being beyond demographic variables. Using nested 10-fold cross-validation, we compared a full model (demographic + driving variables) with a reduced model (demographic variables only). In each fold, Ridge regression was tuned on the training set and evaluated on the test set. Prediction performance was measured using participant-level squared error,  $(y - \hat{y})^2$ . We calculated the difference in squared errors between the full and reduced models, and used a one-sample  $t$ -test to determine whether the full model provided a significant improvement.

The results are summarized in Figure 4. The largest improvement in predictive performance was observed for physical decline (difference in mean squared error ( $\Delta\text{MSE}$ ) =  $-0.06$ ,  $t(2657) = -5.06$ ,  $p < .001$ , Cohen's  $d = -0.10$ ), followed by life satisfaction ( $\Delta\text{MSE} = -0.04$ ,  $t(2657) = -4.87$ ,  $p < .001$ ,  $d = -0.10$ ), role constraints ( $\Delta\text{MSE} = -0.02$ ,  $t(2657) = -3.14$ ,  $p = .01$ ,  $d = -0.06$ ), and fatigue ( $\Delta\text{MSE} = -0.02$ ,  $t(2657) = -2.95$ ,  $p = .02$ ,  $d = -0.06$ ). For all other domains, the full model did not significantly improve out-of-sample predictive performance after Bonferroni correction ( $ps > .13$ ; see Table S4 for full results).

**Figure 3**

*Improvement in Out-of-Sample Prediction for a Full Model Against a Reduced Model.*



*Note.* The full model included all driving variables and all demographic variables. The reduced model only included all demographic variables. The prediction error for each model was quantified as participant-level squared error,  $(y - \hat{y})^2$ . The plot here shows the difference (Full – Reduced) in prediction error for each health variable. Error bars show 95% confidence intervals. N = 2658.

## Discussion

Using one of the largest naturalistic driving datasets to date, we found that everyday driving behaviors are associated with both overall well-being and specific health domains in older adults. Certain driving patterns were uniquely linked to particular domains, and including driving variables significantly improved out-of-sample prediction for several health outcomes. Together, these findings extend existing work that has largely focused on smartphone data and position driving behavior as a potential new source of personal sensing.

One notable pattern was that overall life satisfaction was positively associated with morning rush hour travel but negatively associated with evening rush hour travel. These relationships could reflect multiple underlying mechanisms. Although speculative, one possibility is that morning travel among older adults often involves socially engaging and purposeful activities (e.g., work, volunteering, or group gatherings) that promote a sense of meaning and connectedness. In contrast, evening travel for older adults may be more challenging due to increased fatigue and reduced visibility<sup>28</sup>. Relatedly, role constraints (i.e., perceived difficulty in performing one's usual roles and activities, such as working and hanging out with family and friends) were positively associated with the percentage of evening rush hour travel but negatively associated with the percentage of morning rush hour travel. These results suggest that evening travels for older adults may reflect

obligations rather than choice, such that these travels might displace leisure or social activities and diminish well-being.

Among the specific health domains examined, physical decline stood out as having the strongest associations with driving behaviors. We identified two markers of physical decline: minutes per trip and minutes per chain. Physical decline also showed the largest improvement in out-of-sample predictive performance, indicating that driving data are especially informative for understanding physical functioning. These findings align with prior research showing that declining physical ability is strongly associated with driving reduction and cessation<sup>29</sup>. The two physical decline-related driving signatures suggest that individuals who spend more time per trip or chain tend to have poorer physical functioning. Longer trip or chain durations may result from driving more slowly, accelerating less forcefully, or spending extra time at turns and intersections<sup>30,31</sup>. These changes could reflect strategic adaptations to maintain safety or unintended consequences of declined physical functioning.

Higher travel speed was associated with multiple positive outcomes, including higher life satisfaction and lower levels of physical decline, fatigue, role constraints, depression, and anxiety. Driving at a higher speed requires faster reaction and greater strength to control the vehicle, which are indicative of good physical health. In addition, driving at faster speeds reduces minutes per trip/chain, enables individuals to reach more destinations more efficiently, and therefore grants them access to more opportunities and resources. It is important to note that the average travel speed in our sample was relatively modest ( $M = 26$  mph,  $SD = 6.46$ ; see Figure S1 for the full distribution). Thus, these associations do not necessarily suggest that driving at excessive speeds is associated with better health outcomes, nor do they suggest that driving faster has a causal effect on health outcomes.

Many different types of sensors can be installed in vehicles to capture driving behaviors at varying levels of granularity. Here, we used a GPS device connected to the vehicle's diagnostic port, which primarily records everyday mobility patterns. These measures may be more sensitive to certain aspects of health (e.g., physical functioning) than others (e.g., emotional states). Other sensors, such as in-vehicle and external cameras<sup>22</sup>, can capture fine-grained driving behaviors, including distracted driving (e.g., texting while driving), aggressive driving (e.g., tailgating), driving errors (e.g., lane-merging mistakes), and traffic violations (e.g., failing to yield). Many modern vehicles are also equipped with driver state monitoring systems that can track eye movements, facial expressions, and other physiological signals, which could be leveraged to understand drivers' moment-to-moment mental states<sup>32</sup>. With the increasing digitalization and interconnectedness of mobility systems, these diverse data streams collectively generate vast amounts of data, providing unprecedented opportunities to understand how humans interact with mobility systems reflects their health and well-being.

Several caveats should be noted. First, although many of the relationships observed were statistically significant, the effect sizes were relatively small ( $\beta \approx .10$  in standardized

regression coefficients). Similarly, while models including driving variables significantly improved out-of-sample predictive performance for certain health domains, the magnitude of improvement is not yet sufficient for immediate practical application. The driving measures available in the LongROAD dataset were pre-determined and not tailored to the specific health outcomes examined here, likely underestimating the true associations. With more targeted study planning and feature engineering, it is possible to distill more informative behavioral markers of health and well-being. Second, although older adults represent an important group that tends to exhibit elevated health concerns, our sample consisted exclusively of individuals in this age range, potentially limiting the generalizability of the findings to other age groups. Future research should include a broader range of drivers to better capture age-related differences in driving behavior and health. Together, these limitations highlight the need for continued refinement in both data collection and analysis to realize the potential of naturalistic driving data as a tool for personal sensing.

Naturalistic driving data hold great promise for understanding and predicting health and well-being. Combining these data with other sensing modalities, such as smartphones, could provide a more comprehensive picture of individuals' daily lives. Realizing this potential, however, will require rigorous attention to privacy and ethical issues to ensure that these rich data streams are used responsibly to improve public health.

## **Methods**

### **Participants**

Data came from the AAA Longitudinal Research on Aging Drivers (AAA LongROAD) study<sup>21</sup>. The AAA LongROAD study is a large-scale prospective cohort study seeking to understand older drivers' mobility patterns and health outcomes. Participants were recruited from five study sites in the US: Ann Arbor, Michigan; Baltimore, Maryland; Cooperstown, New York; Denver, Colorado; and La Jolla, California. Eligibility criteria include (1) 65-79 years old at the time of enrollment with a valid driver license, (2) driving on average at least once a week, (3) residing in the study site area for at least 10 months a year, (4) having no plan to move outside of the study site area within the next 5 years, (5) having access to motor vehicle of model year 1996 or newer with an accessible port to install the DataLogger, (6) driving one vehicle more than 80% of the time if having access to more than one vehicle, (7) being fluent in English, (8) without significant cognitive impairment. The Institutional Review Board approved the study's protocol at each site. Detailed descriptions of the LongROAD study's protocol can be found elsewhere<sup>21</sup>. A total of 2990 participants were recruited. After data cleaning (detailed in the "Data Cleaning" section), the final sample consisted of 2658 participants.

### **Sensitivity Power Analysis**



We conducted a sensitivity power analysis to examine the minimum effect size that can be reliably detected given the study's final sample size. For the bivariate analysis, power was estimated based on a multiple regression examining the bivariate relationship between a driving variable and an outcome variable while controlling for demographic variables. This analysis shows that a standardized regression coefficient of 0.054 can be detected with 80% power in a two-tailed *t*-test with alpha of 0.05.

For the multivariate analysis, power was estimated in a simplified path model with one driving variable predicting two outcome variables. We compared a model where the two paths were free against an alternative model where the two paths were constrained to be equal. The analysis shows that a slope difference (in terms of a difference in standardized regression coefficients) of 0.077 can be detected with 80% power with an alpha of 0.05.

Finally, for a one-sample *t*-test, the current sample can reliably detect an effect size (in terms of Cohen's *d*) of 0.048 with 80% power.

## **Health and Well-Being Measures**

The study consists of a baseline session (Year 0) and four annual follow-up sessions (Years 1 – 4), with sessions roughly 12 months apart. During the baseline and each follow-up session, participants completed, among other measures, questions about life satisfaction and the PROMIS (Patient-Reported Outcomes Measurement Information System) short-form version<sup>33</sup>. PROMIS is a system of measures for assessing cognitive, physical, social, and mental health that is widely used in both clinical and non-clinical populations. All questions used in this study can be found in the supplemental material.

Cognitive health was measured by five items assessing perception of cognitive issues in areas such as concentration, memory, and mental acuity (e.g., "In the past 7 days, my thinking has been slow.").

Physical health included assessments of physical decline and fatigue. Physical decline was measured by four items assessing perception of physical issues such as having difficulty doing chores, going up and down stairs, walking, and running errands. (e.g., "Are you able to go up and down stairs at a normal pace?"). Fatigue was measured by four items assessing feelings of tiredness and exhaustion (e.g., "In the past 7 days, I feel fatigued.").

Social health included assessments of role constraints, social isolation, informational support, emotional support, and instrumental support. Role constraints were measured by four items assessing perceived difficulty in performing one's usual roles and activities (e.g., "I have trouble doing all of my regular leisure activities with others."). Social isolation was measured by four items assessing perceptions of being isolated, excluded, disconnected from, or unknown by others (e.g., "I feel left out."). Informational

support was measured by four items assessing perceived availability of helpful information or advice when needed (e.g., “I have someone to give me good advice about a crisis if I need it.”). Emotional support was measured by four items assessing feelings of being cared for and valued as a person and having confident relationships (e.g., “I have someone who will listen to me when I need to talk.”). Instrumental support was measured by four items assessing perceived availability of material, task, or medical assistance (e.g., “Do you have someone to help you if you are confined to bed?”).

Mental health included assessments of depression, anxiety, and anger. Depression was measured by four items assessing feelings of hopelessness, worthlessness, helplessness, or sadness (e.g., “In the past 7 days, I felt depressed.”). Anxiety was measured by four items assessing feelings of anxiety, fear, or being overwhelmed by worry (e.g., “In the past 7 days, my worries overwhelmed me.”). Anger was measured by five items assessing feelings of anger, annoyance, grouching, or being ready to explode (e.g., “In the past 7 days, I felt angry.”).

Finally, overall life satisfaction was measured by five items assessing the extent to which participants were satisfied with their (1) daily life and leisure activities, (2) family life, (3) financial situation, (4) total household income, and (5) health.

All items were rated on a 1-to-5 scale. The PROMIS health measures were computed as average scores and reverse-coded when necessary, such that higher scores indicated a worse condition in the measured construct (e.g., higher scores reflected higher levels of role constraints). Life satisfaction was computed as an average score and not reverse-coded. Thus, a higher score indicated a higher level of life satisfaction. All measures' internal reliability (in terms of Cronbach's  $\alpha$  at baseline) ranged from .76 to .90.

## **Driving Measures**

During participants' baseline visit, a data recording device, DataLogger (Danlaw, Inc.), was installed in their primary vehicle's on-board diagnostic port. The DataLogger collected vehicle speed and GPS data for each trip. A trip was defined as ignition-on to ignition-off for events in which the vehicle moved more than 500 feet. To identify who was driving the vehicle, the participant and all other regular drivers of the vehicle were asked to carry a Bluetooth low-energy beacon that broadcast a unique identifier. Driver identity was determined based on the unique identifier. In case multiple identifiers were detected, driver identity was determined based on relative signal strength (the one closest to the DataLogger was determined to be the driver). Trips performed by individuals other than the participant were discarded. Raw data was sent to the LongROAD data center via cellular network at the end of each trip and screened daily by research staff to ensure the DataLogger worked properly. In case of data logging issues, research staff contacted the appropriate site coordinators to investigate further. Once the issue was resolved, the database was corrected accordingly. For example, if the participant reported that they

forgot to wear the Bluetooth beacon during the past week but were still driving, those trips were retained in the database as participant trips.

The raw data were used to create a number of derived variables for each calendar month. Table S1 lists the variables and their definitions used in the current study.

## **Data Cleaning**

During the course of the study, the DataLogger was replaced by a new smartphone-based method because the cellular service to transmit DataLogger data was no longer supported. The new method differed in several technical aspects of data collection. Most participants had already switched to the new method by their third follow-up session, resulting in the majority (96.84%) of DataLogger data being collected before this point. Therefore, the current study analyzed only the DataLogger data recorded before the third follow-up, covering study years 0 - 2. Correspondingly, self-reported data from the same time frame were used for each participant.

Monthly driving data were excluded if the DataLogger was inactive for the entire month. All monthly data in a study year were excluded if that year contained fewer than three full months of driving data. These exclusions removed 8.75% of all DataLogger data. Additionally, values exceeding 3.5 standard deviations from the mean (by study site and study year) were removed for each driving variable.

A median of 29 full months of driving data was available for each participant (range: 3–38 months). 71.67% of participants completed three assessment sessions (baseline and two annual follow-ups), 19.68% completed two sessions, and 8.65% completed only the baseline session. Participant-level means were calculated for each driving behavior and health measure by averaging driving data across months and questionnaire data across test sessions. Listwise deletion was applied for any missing data in driving variables, health measures, or demographic information. This resulted in the exclusion of 332 participants, yielding a final sample size of 2,658.

## **Statistical Analysis**

Statistical analyses were conducted in the R environment. In all analyses, driving variables and health variables were treated as continuous. Geographic location, sex, race, work status, and marital status were treated as factors and dummy-coded. Age, education level, and income level were treated as continuous. Continuous variables were z-scored to obtain standardized regression coefficients. *P*-values were adjusted using the Bonferroni method.

For the bivariate analysis, we used a driving variable to predict an outcome variable while controlling for demographic variables, including geographic location, age, sex, race, home type, work status, marital status, education level, and household income level. The

effect of interest was whether the driving variable predicted the outcome variable over and above demographic variables. Statistical significance was based on two-tailed *t*-tests.

Multivariate analysis was conducted using the R package *lavaan*<sup>34</sup>. We fit a path model that simultaneously includes all bivariate relationships between a given driving variable and all outcome variables while controlling for the same demographic variables. These regression paths reproduce the bivariate analysis results. Importantly, defined parameters were added to compare a “candidate” slope against all other slopes. The candidate slope was the significant bivariate relationship in the bivariate analysis if the driving variable significantly predicted a single outcome measure. If a driving variable significantly predicted multiple outcomes, the strongest association was chosen as the candidate slope. The effects of interest were whether the candidate slope significantly differed from other slopes. All models were fit using the maximum likelihood estimator. Statistical significance was based on two-tailed Wald *z*-tests.

To compare out-of-sample predictions between a reduced model and a full model, we implemented nested *k*-fold cross-validation with Ridge regression. The reduced model included only demographic variables as predictors, whereas the full model additionally included all driving variables. Ridge regression was used because it helps reduce model complexity and prevents overfitting. We used a 10-fold outer loop to partition the data into training and test sets. Within each training set, an inner leave-one-out cross-validation was used to tune the hyperparameter  $\alpha$  (regularization strength) for the Ridge regression. The tuned model was then evaluated on the corresponding test set. This nested cross-validation procedure was applied to both the reduced and full models. Participant-level prediction error was quantified as the squared error,  $(y - \hat{y})^2$ . For each individual, we calculated the difference in squared error between the full and reduced models. A one-tailed one-sample *t*-test was used to determine whether the full model significantly improved predictive performance compared to the reduced model. We used Cohen’s *d* for effect size, calculated as the mean of the differences in prediction error divided by their standard deviation.

### **Data Availability Statement**

The “minimum dataset” that is necessary to interpret, verify, and extend the research in the article can be accessed at <https://github.com/HanZhang-psych/Understanding-Health-and-Well-Being-from-Naturalistic-Driving-Behavior>. Please note that data-sharing of participant-level data is subject to controlled access (see below). However, the results can be reproduced from a correlation matrix and descriptive statistics of the variables, without needing the participant-level data. We provide detailed instructions in the online repository to ensure reproducibility.

Sharing of participant-level data of the AAA LongROAD study with external parties is limited by the consent forms signed by study participants and the human subjects research protection policies of the individual study sites. Restrictions apply to the availability of

these data. Data may be available from David W. Eby (eby@umich.edu), with permission from the AAA Foundation for Traffic Safety and upon execution of a Data Use Agreement, with limitations on use. We strive to respond within a week of initial request. Please refer to <http://aaafoundation.org/wp-content/uploads/2024/02/202402-AAAFTS-LongROAD-External-Data-Sharing-Guidelines.pdf> for more information.

### Code Availability

Analysis code can be accessed at <https://github.com/HanZhang-psych/Understanding-Health-and-Well-Being-from-Naturalistic-Driving-Behavior>.

### References

1. Mohr, D. C., Zhang, M. & Schueller, S. M. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annu. Rev. Clin. Psychol.* **13**, 23–47 (2017).
2. Cornet, V. P. & Holden, R. J. Systematic review of smartphone-based passive sensing for health and wellbeing. *J. Biomed. Inform.* **77**, 120–132 (2018).
3. Kingstone, A., Smilek, D. & Eastwood, J. D. Cognitive Ethology: A new approach for studying human cognition. *Br. J. Psychol.* **99**, 317–340 (2008).
4. Shiffman, S., Stone, A. A. & Hufford, M. R. Ecological Momentary Assessment. *Annu. Rev. Clin. Psychol.* **4**, 1–32 (2008).
5. Insel, T. R. Digital Phenotyping: Technology for a New Science of Behavior. *JAMA* **318**, 1215–1216 (2017).
6. Wright, A. G. C., Ringwald, W. R. & Zimmermann, J. Measuring Psychopathology in Daily Life. *Clin. Psychol. Sci.*
7. Onnela, J.-P. & Rauch, S. L. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology* **41**, 1691–1696 (2016).
8. Trifan, A., Oliveira, M. & Oliveira, J. L. Passive Sensing of Health Outcomes Through Smartphones: Systematic Review of Current Solutions and Possible Limitations. *JMIR MHealth UHealth* **7**, e12649 (2019).
9. Rabbi, M., Pfammatter, A., Zhang, M., Spring, B. & Choudhury, T. Automated Personalized Feedback for Physical Activity and Dietary Behavior Change With Mobile Phones: A Randomized Controlled Trial on Adults. *JMIR MHealth UHealth* **3**, e4160 (2015).
10. Zhang, H., Ibrahim, A., Parsia, B., Poliakoff, E. & Harper, S. Passive social sensing with smartphones: a systematic review. *Computing* **105**, 29–51 (2023).
11. Wang, W. et al. Social Sensing: Assessing Social Functioning of Patients Living with Schizophrenia using Mobile Phone Sensing. in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* 1–15 (ACM, Honolulu HI USA, 2020). doi:10.1145/3313831.3376855.

12. Bufano, P., Laurino, M., Said, S., Tognetti, A. & Menicucci, D. Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *J. Med. Internet Res.* **25**, e46778 (2023).
13. Ringwald, W. R., King, G., Vize, C. E. & Wright, A. G. C. Passive Smartphone Sensors for Detecting Psychopathology. *JAMA Netw. Open* **8**, e2519047 (2025).
14. Zhang, X. & Steinbach, R. *American Driving Survey: 2023*.  
<https://aaaafoundation.org/wp-content/uploads/2024/08/202408-AAAFTS-American-Driving-Survey-2023.pdf> (2024).
15. Michon, J. A. A Critical View of Driver Behavior Models: What Do We Know, What Should We Do? in *Human Behavior and Traffic Safety* (eds Evans, L. & Schwing, R. C.) 485–524 (Springer US, Boston, MA, 1985). doi:10.1007/978-1-4613-2173-6\_19.
16. Eby, D. W., Molnar, L. J. & Louis, R. M. S. The challenges. in *Perspectives and Strategies for Promoting Safe Transportation Among Older Adults* 3–21 (Elsevier, Cambridge, MA, 2019).
17. Cunningham, M. L. & Regan, M. A. The impact of emotion, life stress and mental health issues on driving performance and safety. *Road Transp. Res. J. Aust. N. Z. Res. Pract.* **25**, 40–50 (2016).
18. Jerome, L., Segal, A. & Habinski, L. What We Know About ADHD and Driving Risk: A Literature Review, Meta-Analysis and Critique. *J. Can. Acad. Child Adolesc. Psychiatry* **15**, 105–125 (2006).
19. Fernandes, R., Job, R. F. S. & Hatfield, J. A challenge to the assumed generalizability of prediction and countermeasure for risky driving: Different factors predict different risky driving behaviors. *J. Safety Res.* **38**, 59–70 (2007).
20. Taubman – Ben-Ari, O., Eherenfreund – Hager, A. & Prato, C. G. The value of self-report measures as indicators of driving behaviors among young drivers. *Transp. Res. Part F Traffic Psychol. Behav.* **39**, 33–42 (2016).
21. Li, G. et al. Longitudinal Research on Aging Drivers (LongROAD): study design and methods. *Inj. Epidemiol.* **4**, 22 (2017).
22. Dingus, T. A. et al. Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proc. Natl. Acad. Sci.* **113**, 2636–2641 (2016).
23. Blake, M. et al. A combined naturalistic driving, clinical, and neurobehavioral data set for investigating aging and dementia. *Sci. Data* **12**, 1209 (2025).
24. Di, X. et al. Detecting mild cognitive impairment and dementia in older adults using naturalistic driving data and interaction-based classification from influence score. *Artif. Intell. Med.* **138**, 102510 (2023).
25. Bayat, S. et al. GPS driving: a digital biomarker for preclinical Alzheimer disease. *Alzheimers Res. Ther.* **13**, 115 (2021).
26. Chen, C. et al. Identifying major depressive disorder in older adults through naturalistic driving behaviors and machine learning. *Npj Digit. Med.* **8**, 102 (2025).
27. Carr, D. B. & Babulal, G. M. Addressing the Complex Driving Needs of an Aging Population. *JAMA* **330**, 1187–1188 (2023).
28. Molnar, L. J. & Eby, D. W. The Relationship between Self-Regulation and Driving-Related Abilities in Older Drivers: An Exploratory Study. *Traffic Inj. Prev.* **9**, 314–319 (2008).

29. Edwards, J. D., Lunsman, M., Perkins, M., Rebok, G. W. & Roth, D. L. Driving Cessation and Health Trajectories in Older Adults. *J. Gerontol. Ser. A* **64A**, 1290–1295 (2009).
30. Guo, F., Ai, Y. & Qu, S. Intersection challenges for older drivers: The impact of aging on visual cognition and driving efficiency at crossroads. *Traffic Inj. Prev.* **0**, 1–10.
31. Robertsen, R., Lorås, H. W., Polman, R., Simsekoglu, O. & Sigmundsson, H. Aging and Driving: A Comparison of Driving Performance Between Older and Younger Drivers in an On-Road Driving Test. *Sage Open* **12**, 21582440221096133 (2022).
32. Koesdwiady, A., Soua, R., Karray, F. & Kamel, M. S. Recent Trends in Driver Safety Monitoring Systems: State of the Art and Challenges. *IEEE Trans. Veh. Technol.* **66**, 4550–4563 (2017).
33. Cella, D. *et al.* The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J. Clin. Epidemiol.* **63**, 1179–1194 (2010).
34. Rosseel, Y. lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.* **48**, 1–36 (2012).