Examining urban and rural bicycling in the United States

Calvin P. Tribby and Doug S. Tharp

EXECUTIVE SUMMARY

Introduction

Bicycling is an uncommon travel behavior in the United States, yet it has personal health and community economic benefits (Krizec 2007; Dill 2009). There have been large increases in bicycling use since 2001 through 2016, although overall rates remain low at 0.6% of US commuters (Pucher et al. 2011; Alliance for Biking & Walking 2016). There has been considerable investment in the provision of bicycling infrastructure, such as bicycle lanes and shared bicycle systems, which provide safe and convenient environments to encourage more bicycling (Dill 2009; Faghih-Imani et al. 2014; Hirsch et al. 2017). However, most of the improvements to the bicycling infrastructure are concentrated in urban areas and less is known about bicycling in the rural context. This research project estimates the urban and rural prevalence of shared use bicycling and bicycling for exercise in the US. This research also explores the socio-demographic, environmental, and perceptual correlates of bicycling use.

Insight into bicycling by considering urban and rural contexts is important, as rural environments are increasingly being viewed as places of health disparities (Blake et al. 2017; Wheeler and Davis 2017). Estimating the prevalence of rural cycling is important to determining if policy and built environment improvements in rural areas are needed to increase bicycling for improving physical activity levels and health. Furthermore, bicycle-related infrastructure in rural areas is a potential economic development strategy to support local businesses (Bowker, Bergstrom, and Gill 2007).

Research Questions

This first research question is: what is the prevalence of shared bicycle use and bicycling for exercise? This question assesses both the independent and joint prevalence of shared bicycle use and exercise bicycling for urban and rural contexts for the US overall and within Census divisions. The second research question is: what are the socio-demographic, environmental, and perceptual correlates of bicycle use. This question uses a machine learning technique to identify important variables to distinguish between people who do not bicycle and those that do and we also look at rural only residents in a separate analysis.

Methodology

Sample. We used the 2015 National Household Travel Survey (NHTS) Person File, sponsored by the Federal Highway Administration and administered by Westat. The NHTS is a household survey of a nationally representative sample of the US non-institutionalized population (age ≥5 years) with sample design and data collection details available online (Federal Highway Administration 2018). We used the combined national sample and state add-on samples, for a total sample of adults interviewed of 264,234 persons. We excluded respondents who had medical conditions affecting travel, those missing nativity (born in US) status, missing housing unit density, proportion renter occupied housing, and set other variables as missing if response was “I don’t know,” “I prefer not to answer,” or was an appropriate skip due to a response to a previous question. The analytic sample included 236,959 persons.

Measures. The bicycling variables used were count of bicycling trips for exercise and the count of shared bicycle use in the past 7 days. These two variables were recoded to any or no trips and a third variable was created based on their 4 unique combinations (no bicycling, bicycling for exercise only, shared bicycle use only, or both bicycling for exercise and shared use). The urban/rural variable was defined as the classification of the respondent’s residential address county based on the US Census definition. The geographic variable is Census divisions and there are nine divisions. Details of survey responses by division is in the Appendix, Table A1. The complete list of independent variables is in Table A2.

Statistical Analysis. To estimate the prevalence of bicycling use in the US, we used survey procedures (proc surveyfreq) to estimate the frequencies of bicycling. This included survey weights and jackknife variance estimates for 95% confidence intervals, provided in the person weights file. We used SAS v9.4 (Cary, NC).

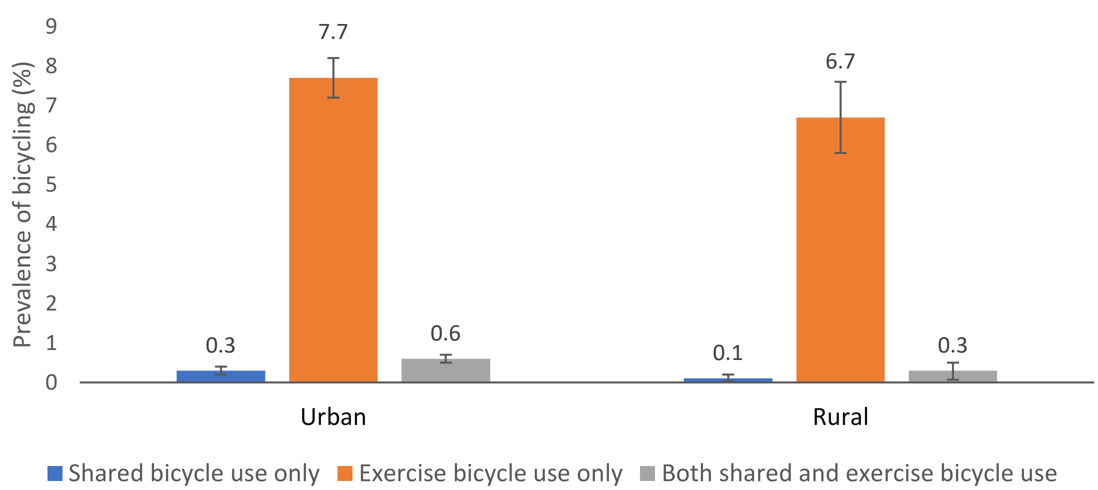
To model the association between bicycling and covariates, we used random forest. Traditional statistical methods have difficulty with large attribute datasets, including overfitting, multicollinearity and sensitivity to outliers. The basis of the random forest algorithm is decision trees, which is a method that successively uses sets of explanatory variables to best predict an outcome variable (Breiman 2001). The random forest algorithm creates a large collection of decision trees using a random subset of variables from the entire independent variable set to create each tree. The random forest splits the input data into two sets: the training data used to create the trees and the test data used for determining model accuracy and variable importance. We implemented the random forest models with the randomForest v4.6-14 package in R v3.4.2. The key parameters to specify for the random forest model are: mtry, the number of variables to sample at each split point; ntree, the number of trees to grow; nodesize, the minimum size of the terminal nodes; and, sampsize, the size of the sample to draw. We set ntree=100 and left the other parameters at their defaults (mtry=7; nodesize=1; sampsize= 62921).

We used the out-of-bag (OOB) performance estimates for ranking variable importance. The OOB estimates use a subset of the input data, also called the test dataset. We also used partial dependence plots, which are the marginal effect of each variable (the logits) for the classification of persons as a bicyclist. The less negative the marginal effect, the better the given value of the independent variable for classifying persons as bicyclists.

Results

The overall prevalence of urban shared bicycling and exercise cycling in the US was 8.5% (95% CI: 8.1%-9.0%). In rural areas, the bicycling prevalence was 7.1% (95% CI: 6.3%-8.0%). Full results in Appendix Table A2. Figure 1 shows the estimated prevalence of bicycling for exercise only in urban areas (7.7%, 95% CI: 7.2%-8.2%) and rural areas (6.7%, 95% CI: 5.8-7.6%). Shared bicycle use only and both shared and exercise bicycling was about double in urban areas compared to rural areas (Figure 1).

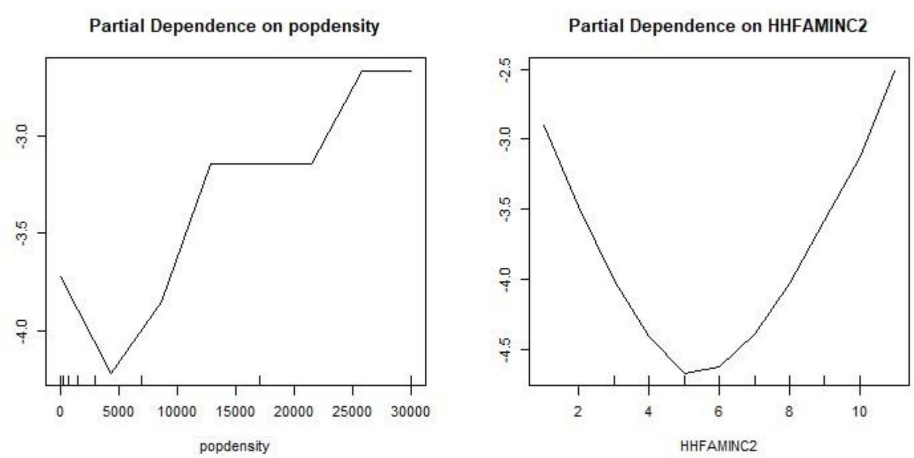
Figure 1. Weighted prevalence (95% CI) of any shared and/or exercise cycling trips in the past 7 days by urban/rural status of the non-institutional US population (ages 5+), 2017 National Household Travel Survey



The distribution of bicycle use by Census division indicated that the highest shared bicycle use only (0.7%, 95% CI: 0.5%-0.9%) was in urban areas in the Mid Atlantic (New York, New Jersey, and Pennsylvania), whereas the highest bicycling for exercise only (10.7%, 5.9%-15.6%) was in rural areas in the Mountain division (Arizona, Colorado, Idaho, Montana, New Mexico, Nevada, Utah, and Wyoming) (Appendix, Table A2).

The ten most important variables for classifying persons as bicyclists versus non-bicyclists were: 1) population density; 2) household count of different travel modes; 3) household income; 4) taxi use; 5) household count of vehicles; 6) household count of trips; 7); age 8) household size; 9) smartphone; and, 10) worker density (Appendix, Figure 1A). We estimated two models, the first with perceptions of barriers to cycling and the second without. We did this because the perceptions were limited to those who bicycled, so their importance was an artifact of the survey; subsequent results reference the models omitting bicycling environment perceptions, which had an OOB error rate of 8.94%. The nature of the relationship of these variables for classifying persons as bicyclists or non-bicyclists varied. For example, the effect of population density and household income are U-shaped: lower values and higher values better classified persons as bicyclists (Figure 2, full results in Appendix Figure 2A). The nadir for population density was about 5,000 people per square mile and for household family income is the category of \$35,000-\$49,999 for categorizing persons as bicyclists. This suggests low-density or high-density areas had more cyclists, compared to suburban areas. Similarly, for income: persons with lower income or higher income were more likely to bicycle.

Figure 2. Partial dependence plots of the logits of population density (left) and household family income (right) on classification of persons as bicyclists versus non-bicyclists.



For rural residents, the ten most important variables for classifying persons as bicyclists versus non-bicyclists were: 1) age; 2) number of walking trips; 3) walking for exercise; 4) household count of trips; 5) household count of different modes; 6) distance to work or school (whichever is further); 7) household size; 8) household family income; 9) worker density; and, 10) count of public transit usage (results not shown). The OOB error rate was 7.88%.

Discussion and Conclusion

This research estimated the national prevalence of shared use bicycling and bicycling for exercise. We found that urban and rural areas had similar prevalence of bicycling for exercise. Surprisingly, we also found that prevalence of shared bicycle use was similar between persons who lived in urban and rural areas. While shared bicycle systems are part of the urban bicycle infrastructure, these results suggest that these systems benefit users who do not live in urban areas, such as rural residents, in addition to tourists (O'Brien, Cheshire, and Batty 2014).

The random forest analysis of variables for best classifying overall persons as bicycle users versus non-users revealed that population density around the residence was the most important factor, confirming previous research (Saelens, Sallis, and Frank 2003; Moudon et al. 2005; Winters et al. 2010). While previous research found that factors such as higher income were associated with more overall trips (Ortúzar and Willumsen 2001), we are not aware of previous research about household travel mode diversity as a significant predictor of bicycle use. The overall findings contrasted with the result for rural residents that age was the most important variable and count of walking trips was second most important. These results support some previous research that found age influences on bicycling (Hunt and Abraham 2007) and synergies with walking (Pucher et al. 2010). The contrast between overall predictors and rural-specific predictors for classifying bicyclists warrants further investigation.

This research suggests that rural areas have comparable rates of cycling as urban areas. This travel behavior, bicycling, is not likely to lead to the observed health disparities in rural populations. However, this research also suggests factors that are important for rural persons to bicycle are likely different from urban persons. This encourages the largely urban focused bicycle research to attend to rural populations' bicycle needs to promote physical activity and create opportunities for economic development in rural communities.

APPENDIX

References

Alliance for Biking & Walking. 2016. *Bicycling and walking in the United States: 2016 benchmarking report*. Alliance for Biking & Walking Washington, USA.

Blake, K. D., J. L. Moss, A. Gaysynsky, S. Srinivasan, and R. T. Croyle. 2017. Making the Case for Investment in Rural Cancer Control: An Analysis of Rural Cancer Incidence, Mortality, and Funding Trends. *Cancer Epidemiology and Prevention Biomarkers*. http://cebp.aacrjournals.org/content/early/2017/06/09/1055-9965.EPI-17-0092 (last accessed 1 October 2018).

Bowker, J. M., J. C. Bergstrom, and J. Gill. 2007. Estimating the Economic Value and Impacts of Recreational Trails: A Case Study of the Virginia Creeper Rail Trail. *Tourism Economics* 13 (2):241–260.

Breiman, L. 2001. Random Forests. *Machine Learning* 45 (1):5–32.

Dill, J. 2009. Bicycling for Transportation and Health: The Role of Infrastructure. *Journal of Public Health Policy* 30 (S1):S95–S110.

Faghih-Imani, A., N. Eluru, A. M. El-Geneidy, M. Rabbat, and U. Haq. 2014. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography* 41:306–314.

Federal Highway Administration. 2018. National Household Travel Survey. https://nhts.ornl.gov/documentation (last accessed 30 September 2018).

Hirsch, J. A., K. A. Meyer, M. Peterson, L. Zhang, D. A. Rodriguez, and P. Gordon-Larsen. 2017. Municipal investment in off-road trails and changes in bicycle commuting in Minneapolis, Minnesota over 10 years: a longitudinal repeated cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity* 14 (1):21.

Hunt, J. D., and J. E. Abraham. 2007. Influences on bicycle use. *Transportation* 34 (4):453–470.

Krizec, K. J. 2007. Estimating the Economic Benefits of Bicycling and Bicycle Facilities: an Interpretive Review and Proposed Methods. In *Essays on Transport Economics*, Contributions to Economics., eds. P. Coto-Millán and V. Inglada, 219–248. Heidelberg: Physica-Verlag HD https://doi.org/10.1007/978-3-7908-1765-2_14 (last accessed 1 October 2018).

Moudon, A. V., C. Lee, A. D. Cheadle, C. W. Collier, D. Johnson, T. L. Schmid, and R. D. Weather. 2005. Cycling and the built environment, a US perspective. *Transportation Research Part D: Transport and Environment* 10 (3):245–261.

O'Brien, O., J. Cheshire, and M. Batty. 2014. Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography* 34:262–273.

Ortúzar, J. de D., and L. G. Willumsen. 2001. *Modelling transport* 3rd ed. Chichester ; New York: J. Wiley.

Pucher, J., R. Buehler, D. R. Bassett, and A. L. Dannenberg. 2010. Walking and Cycling to Health: A Comparative Analysis of City, State, and International Data. *American Journal of Public Health* 100 (10):1986–1992.

Pucher, J., R. Buehler, D. Merom, and A. Bauman. 2011. Walking and Cycling in the United States, 2001–2009: Evidence From the National Household Travel Surveys. *American Journal of Public Health* 101 (S1):S310–S317.

Saelens, B. E., J. F. Sallis, and L. D. Frank. 2003. Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine* 25 (2):80–91.

Wheeler, S. B., and M. M. Davis. 2017. "Taking the Bull by the Horns": Four Principles to Align Public Health, Primary Care, and Community Efforts to Improve Rural Cancer Control. *The Journal of Rural Health* 33 (4):345–349.

Winters, M., M. Brauer, E. M. Setton, and K. Teschke. 2010. Built Environment Influences on Healthy Transportation Choices: Bicycling versus Driving. *Journal of Urban Health* 87 (6):969–993.

Tables

Table A1. Census division states and sample sizes

Census Divisions	States	Sample (n)	Percent of Total Sample	Percent Urban in Sample
New England	ME, NH, VT, CT, MA, RI	3,932	1.5%	79.7%
Middle Atlantic	NY, NJ, PA	37,837	14.3%	85.5%
East North Central	IL, IN, MI, OH, WI	29,898	11.3%	77.7%
West North Central	IA, KS, MO, MN, ND, NE, SD	10,035	3.8%	72.2%
South Atlantic	DC, DE, FL, GA, MD, NC, SC, WV, VA	57,894	21.9%	80.9%
East South Central	AL, KY, MS, TN	2,641	1.0%	62.8%
West South Central	AR, LA, OK, TX	54,829	20.8%	81.2%
Mountain	AZ, CO, ID, MT, NM, NV, UT, WY	10,386	3.9%	87.7%
Pacific	AK, CA, HI, OR, WA	56,782	21.5%	93.7%

Table A2. Variable abbreviations and descriptions

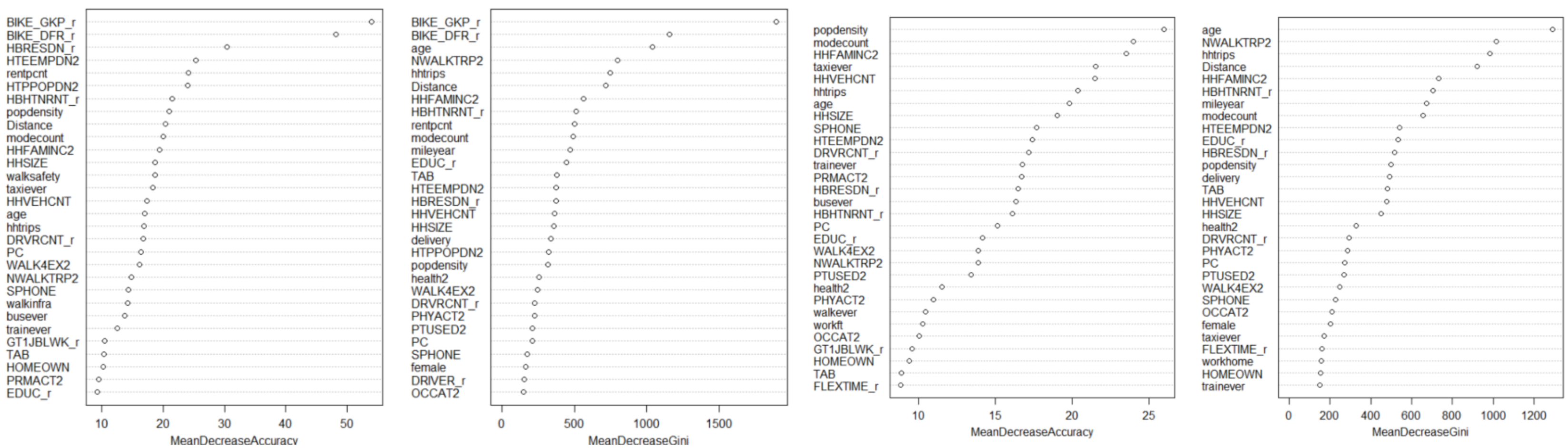
Variable	Description	Variable	Description	Variable	Description	Variable	Description
HOMEOWN	Home ownership	hhtrips	household trips	carever	car use	mileyear	miles travelled per year
HHSIZE	Household size	HBRESDN	housing unit density	walkever	walk use	workhome	ability to work from home
HHVEHCNT	Household vehicle count	mode count	modecount	busever	bus use	workft	full time work
PC	personal computer	active trips	active	taxiever	taxi use	walksafety	perception of safety for walking
SPHONE	smartphone	household active trips	hhactive	paraever	paratransit use	walkinfra	perception of infrastructure for walking
TAB	tablet	delivery	package deliveries	trainever	train use	female	female
popdensity	population density	carshare2	car share use	age	age	white	white, non-Hispanic
PTUSED2	any public transit	children	number of children	hispanic	Hispanic	walk4ex2	walk for exercise
PRMACT2	Primary Activity in Previous Week	HTEEMPDN2	worker density	BORNINUS_r	Nativity	NWALKTRP2	number of walking trips
PHYACT2	physical activity level	HHFAMINC2	household income	DELIVER_r	delivery of packages	driver_r	driving ability
OCCAT2	job category	health2	self-rated health	Distance	distance to work or school, whichever is greater	DRVRCNT_r	count of drivers in household
EDUC_r	education level	FLEXTIME_r	flexible work hours	TravDayHM	travel day start location	GT1JBLWK_r	More than one job

Table A3. Weighted prevalence of any shared and/or exercise cycling trips in the past 7 days by urban/rural status of the non-institutional US population (ages 5+), 2017 National Household Travel Survey

	Urban Prevalence (%) and 95% CI				Rural Prevalence (%) and 95% CI				*P* [a]
	No cycling	Shared Only	Exercise Only	Both Shared and Exercise	No cycling	Shared Only	Exercise Only	Both Shared and Exercise	
US Overall	91.5 (91.0-92.0)		8.5 (8.1-9.0)		92.9 (92.0-93.7)		7.1 (6.3-8.0)		0.01
US Categories	91.5 (91.0-92.0)	0.3 (0.2-0.4)	7.7 (7.2-8.2)	0.6 (0.5-0.6)	92.9 (92.0-93.7)	0.1 (.03-0.2)	6.7 (5.8-7.6)	0.3 (0.07-0.5)	<0.0001
Census Division									
New England	92.6 (90.8-94.4)	0.1 (0.0-0.3)	7.0 (5.2-8.9)	0.3 (0.0-0.6)	92.3 (90.1-94.4)	-	7.3 (5.1-9.4)	0.5 (0.0-1.1)	-
Middle Atlantic	92.2 (90.8-93.5)	0.7 (0.5-0.9)	6.6 (5.3-7.9)	0.6 (0.3-0.8)	93.1 (91.3-94.8)	0.3 (0.0-0.9)	6.5 (5.3-7.8)	0.1 (0.0-0.2)	0.01
East North Central	91.5 (89.5-93.6)	0.3 (0.1-0.6)	7.4 (5.2-9.6)	0.7 (0.3-1.1)	91.8 (89.4-94.2)	0.01 (0.0-0.3)	8.0 (5.6-10.5)	0.2 (0.0-0.4)	0.01
West North Central	90.6 (86.9-94.3)	0.1 (0.0-0.2)	9.1 (5.5-12.8)	0.2 (0.0-0.3)	94.9 (92.6-97.2)	0.01 (0.0-0.03)	5.0 (2.8-7.2)	0.1 (0.0-0.2)	0.06
South Atlantic	90.9 (89.9-91.9)	0.2 (0.1-0.3)	8.3 (7.4-9.2)	0.5 (0.3-0.7)	92.5 (91.1-93.9)	0.2 (0.02-0.4)	6.9 (5.6-8.2)	0.4 (0.1-0.6)	0.36
East South Central	92.4 (89.8-95.1)	0.2 (0.0-0.8)	6.9 (4.4-9.4)	0.4 (0.0-0.9)	96.5 (93.8-99.2)	0.07 (0.0-0.2)	3.0 (0.4-5.7)	0.4 (0.0-1.0)	0.01
West South Central	92.5 (91.6-93.5)	0.1 (0.05-0.2)	6.8 (5.9-7.8)	0.5 (0.4-0.7)	92.7 (91.0-94.4)	0.1 (0.0-0.3)	6.6 (4.8-8.4)	0.6 (0.0-1.4)	0.99
Mountain	89.1 (87.0-91.3)	0.3 (0.1-0.5)	10.0 (8.2-11.9)	0.5 (0.0-1.1)	88.7 (84.0-93.4)	0.04 (0.0-0.1)	10.7 (5.9-15.6)	0.5 (0.0-1.3)	0.03
Pacific	91.6 (90.9-92.3)	0.2 (0.1-0.3)	7.5 (6.8-8.2)	0.7 (0.5-1.0)	90.8 (88.0-93.6)	0.1 (0.0-0.5)	8.9 (6.1-11.8)	0.2 (0.0-0.4)	<0.0001

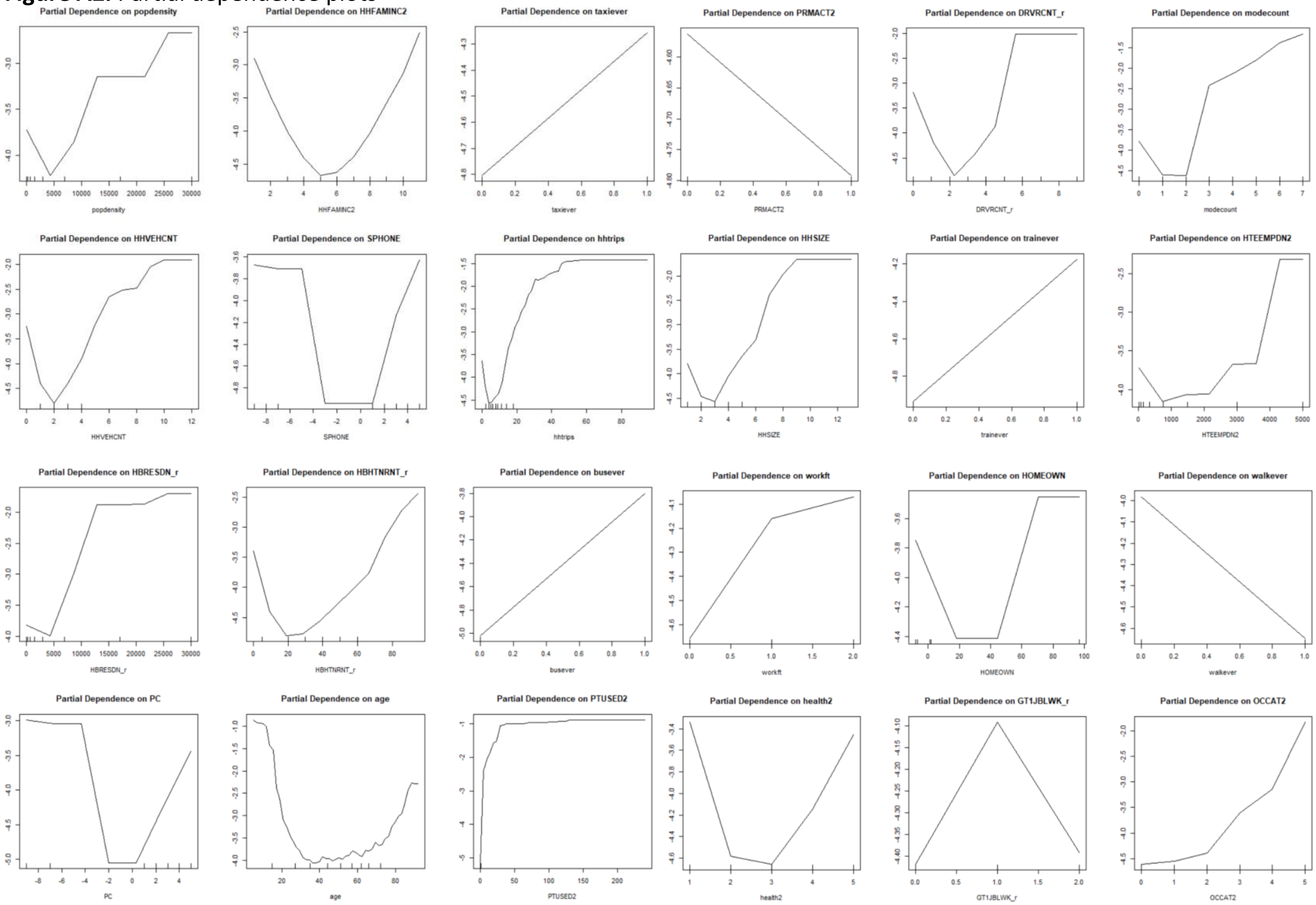
[a]Adjusted Wald *F P*-values

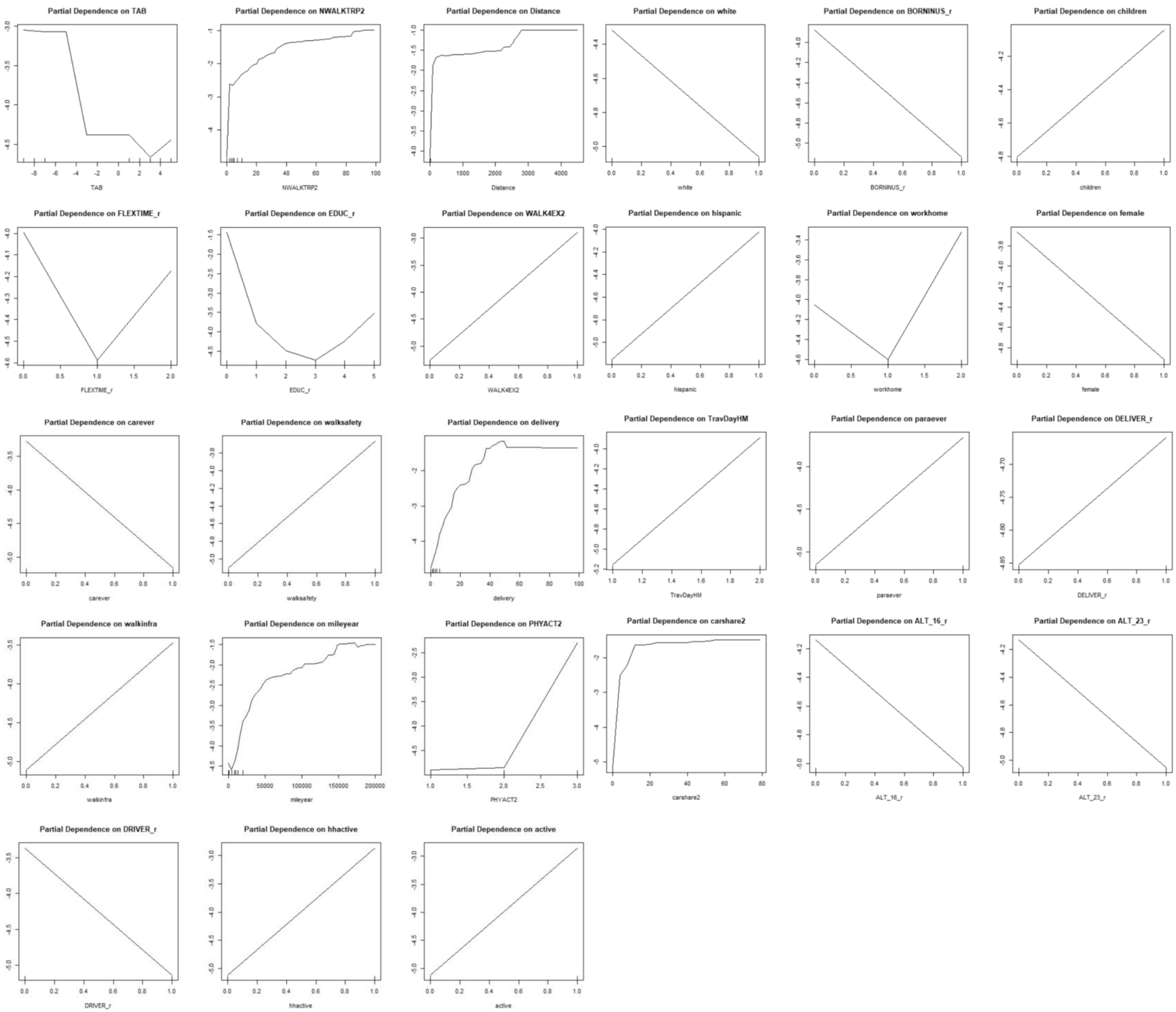
Figure A1. Variable importance plots: A, including bicycling perceptions and B excluding perceptions. A and B mean decrease in accuracy (left) and mean decrease in Gini coefficient (right)



A **B**

Figure A2. Partial dependence plots





Partial Dependence on TAB
Partial Dependence on NWALKTRP2
Partial Dependence on Distance
Partial Dependence on white
Partial Dependence on BORNINUS_r
Partial Dependence on children
Partial Dependence on FLEXTIME_r
Partial Dependence on EDUC_r
Partial Dependence on WALK4EX2
Partial Dependence on hispanic
Partial Dependence on workhome
Partial Dependence on female
Partial Dependence on carever
Partial Dependence on walksafety
Partial Dependence on delivery
Partial Dependence on TravDayHM
Partial Dependence on paraever
Partial Dependence on DELIVER_r
Partial Dependence on walkinfra
Partial Dependence on mileyear
Partial Dependence on PHYACT2
Partial Dependence on carshare2
Partial Dependence on ALT_16_r
Partial Dependence on ALT_23_r
Partial Dependence on DRIVER_r
Partial Dependence on hhactive
Partial Dependence on active