Marginal Effects and Predictive Margins

Choosing and Using a Method of Presenting Estimation Results that Fits Your Research Objectives

6/21/2023

Rebecca A. Raciborski, PhD
Center for Mental Healthcare and Outcomes Research, Central Arkansas Veterans Health System

Poll question 1

- Which of the following have you used to discuss regression results from your research? [select all that apply]
  - ❑ Regression coefficients (incl. standardized coefficients)
  - ❑ Transformed values of coefficients (e.g., odds ratios, incidence rate ratios)
  - ❑ Marginal or incremental effects (also called partial effects)
  - ❑ Predictive margins (also called adjusted predictions or marginal predictions)
  - ❑ None of these

Overview of the talk

- Graphical introduction to marginal effects and predictive margins
  - Simulated data used to illustrate concepts with number of specialty mental health visits as our dependent variable
  - Analyzing a continuous covariate (effect of age)
  - Analyzing a categorical covariate (effect of rural residence)
  - Application of marginal effects to nonlinear models (Poisson for counts)
  - Predicting population average outcomes with predictive margins
- More advanced “choose and use” worked examples with probability of receiving any specialty mental health care as the dependent variable

Suppose we have observational data about receipt of mental health services and related covariates

- subid : Subject ID
- nvisits : Number of specialty mental health care (SMHC) encounters for depression in the last year
- smhc : Binary variable with 1=received any SMHC and 0=received no SMHC
- urban : Binary variable with 1=lives in an urban/suburban ZIP code and 0 if rural
- age : Age in years
- income : Household income in $1,000s
- pvehicle : Binary variable with 1=has a personal vehicle in household & 0 otherwise

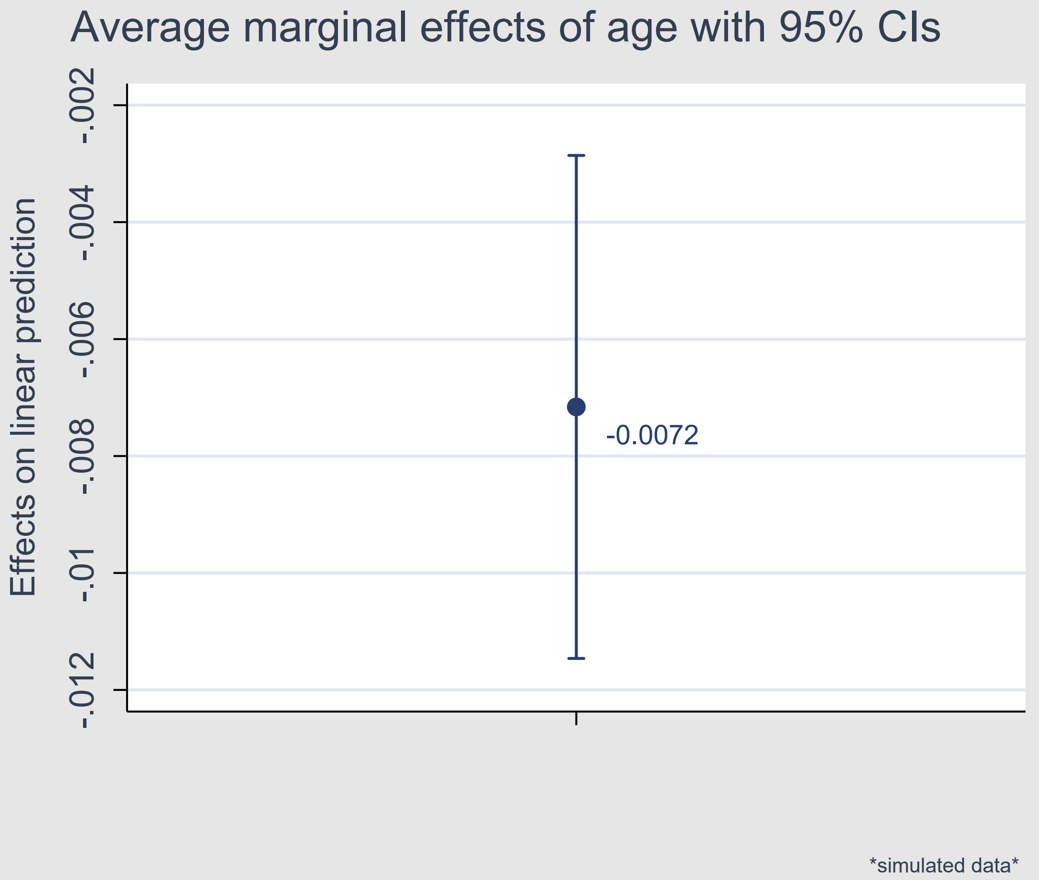
Our variables are distributed like this

	Lives in urban/suburban ZIP		
	No	Yes	Total
N (%)	4,736 23.7%	15,264 76.3%	20,000 100.0%
Number of specialty mental health visits, mean (SD)	1.9 (1.6)	2.8 (2.2)	2.6 (2.1)
Received specialty mental health care			
Yes (%)	82.6%	90.0%	88.3%
Age in years, mean (SD)	45.1 (6.6)	43.7 (6.6)	44.0 (6.7)
Household income in $1,000s, mean (SD)	51.1 (35.8)	162.2 (271.5)	135.9 (242.4)
Personal vehicle in household			
Yes (%)	91.0%	92.8%	92.4%

A marginal effect is, loosely, the expected change in the outcome for a given change in the variable of interest

- In a classic linear model, the *average* marginal (or partial) effects are just the coefficients

$$y = 1.854 - 0.0072 \times age\ in\ years - 0.8935 \times rural\ resident + 1.3644 \times owns\ personal\ vehicle$$

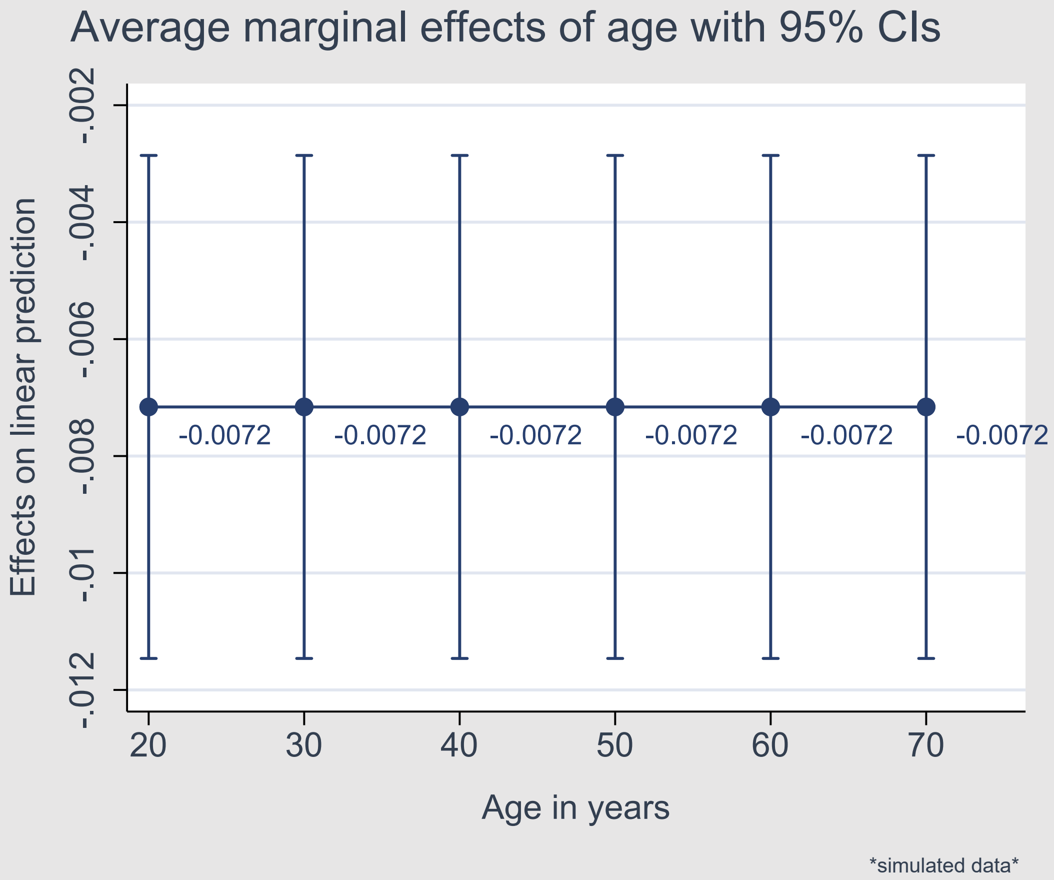


A marginal effect is, loosely, the expected change in the outcome for a given change in the variable of interest

- In a classic linear model, the *average* marginal (or partial) effects are just the coefficients

$$y = 1.854 - 0.0072 \times age\ in\ years - 0.8935 \times rural\ resident + 1.3644 \times owns\ personal\ vehicle$$

- And, with no interactions, the effect is the same for all the values of a continuous covariate



When models include powers & interactions, coefficients do not have a direct marginal effect interpretation

- Suppose our fitted model is instead

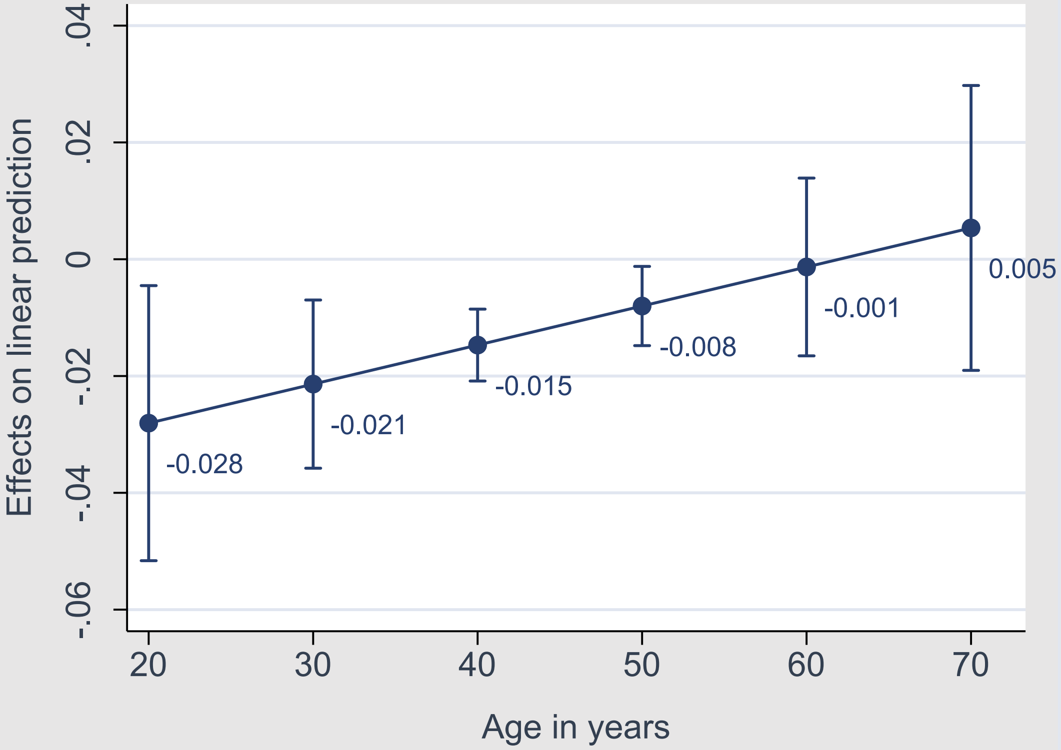
$$y_i = 3.7493 - 0.0414\, age_i + 0.0003\, age_i^2$$

- Different types of marginal effects are obtained by evaluating the derivative of the *conditional expectation*
  - The "conditional expectation" is the predicted value of the outcome given (i.e., conditional on) what we know from the data, denoted E(y | **x**)

$$\frac{d\mathrm{E}(y \mid \boldsymbol{x})}{d\ age} = -0.0414 + (2 \times 0.0003)age$$

- Marginal effects are labeled "conditional marginal effects" here because we specified (i.e., "conditioned on") values of age & there are no other variables in the model
  - Effects as if everyone in the sample was 20, 30, etc.

Conditional marginal effects of age with 95% CIs



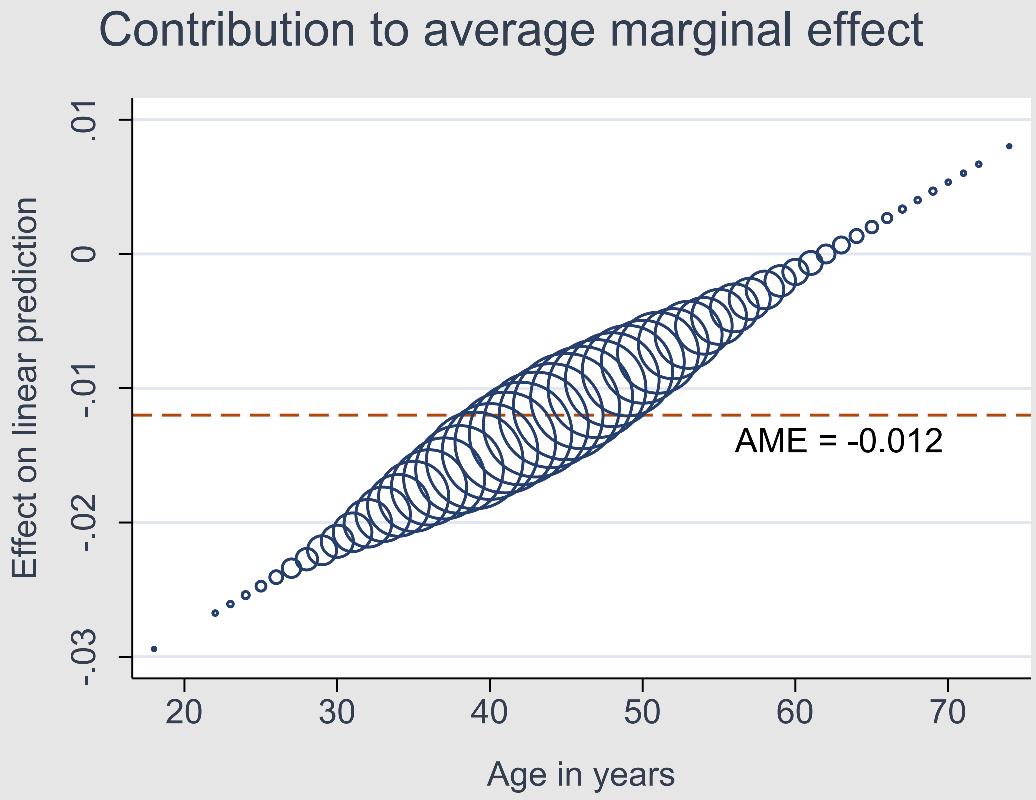
simulated data

The average marginal effect is computed from observation-level evaluation of the derivative

- The *average marginal effect* (AME) of age for this model, using these data is

$$\frac{\sum_{i=1}^{N}\{-0.0414 + 0.0006\ age_i\}}{N} = -0.012$$

- Alternately, we can use the sample average value of age (44 years) to obtain the *marginal effect at the mean* (MEM)
  - In linear models the AME and MEM are equal
- Or select a set of values to represent a "typical" member of the population (e.g., average age of 45 in rural areas) to obtain the *marginal effect at a representative value* (MER)



Size of marker is proportionate to number of observations

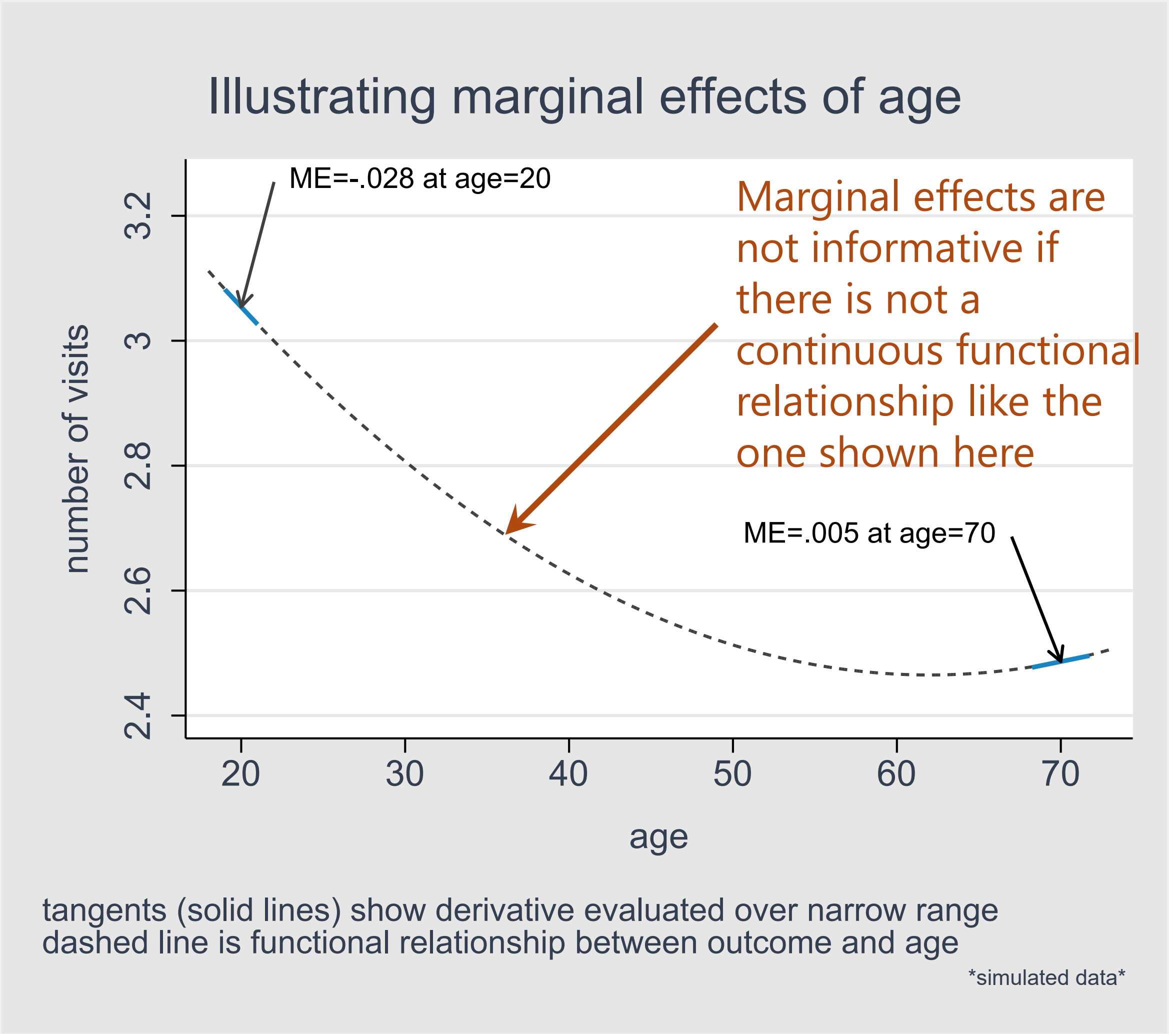
simulated data

The marginal effect captures a continuous rate of change

- The derivative assumes a continuous functional relationship between the expectation and the covariate

$$-0.0414 + 0.0006\,age$$

- This does not align conceptually with variables that only change discretely
  - E.g., living primarily in a USDA-designated rural area vs not
- Nor does it make sense when change is/can only be measured discretely
  - E.g., owning a private vehicle vs proportion of time a private vehicle is available for use

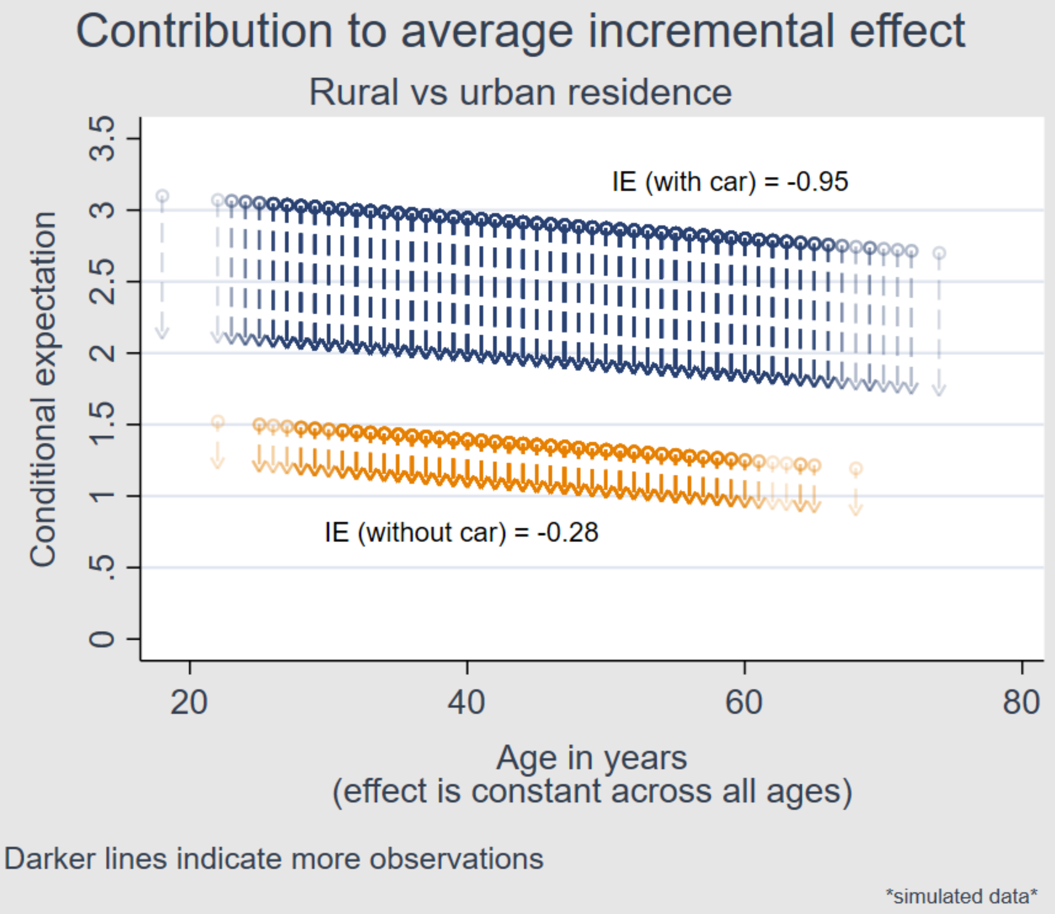


For binary & categorical variables, typically use incremental effects in place of marginal effects

- With no interactions, the incremental effect in a linear model is the coefficient on the indicator for the level
- When there are interactions, incremental effects are obtained by taking differences in conditional expectations
  - Differences are between the expectation, E(y | z=1, **w**), as if everyone in the sample belongs to the category of interest, *z*, and E(y | z=0, **w**) as if everyone in the sample belongs to the base category
  - <u>Not</u> the same as the average outcomes, y, when z=1 vs z=0 in the actual data!
- Just like marginal effects, reported numbers may be a single average incremental effect or computed at representative values (including the mean)

Effect size varies over variables included in the interaction, but is constant across other variables

- Suppose our fitted model is now $y_i = 1.681 - 0.007\, age_i - 0.281\, rural_i + 1.55\, pvehicle_i - 0.67(rural_i \times pvehicle_i)$
- The difference between living in a rural area vs an urban area for those without a car is -0.281 (the effect of rural only)
- The difference for those with a car is $-0.281 - 0.67 = -0.951$
- The average incremental effect of living in a rural vs urban area is about -0.9
  - Recall: >90% of our sample has a car

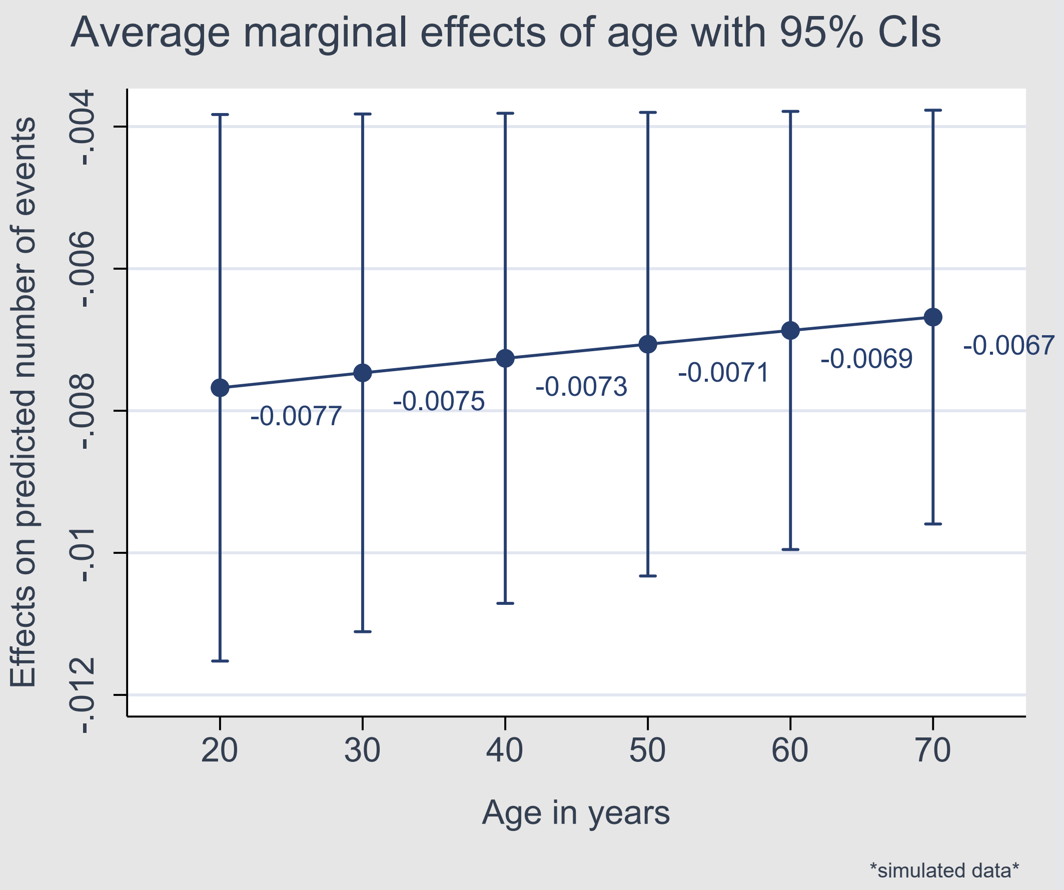


Poll question 2

- When we report a variable’s average marginal or incremental effect, what determines the value we report?
  - The magnitude of the variable's effect
  - The relative proportions of subjects with each possible value of the covariate in the sample
  - Both

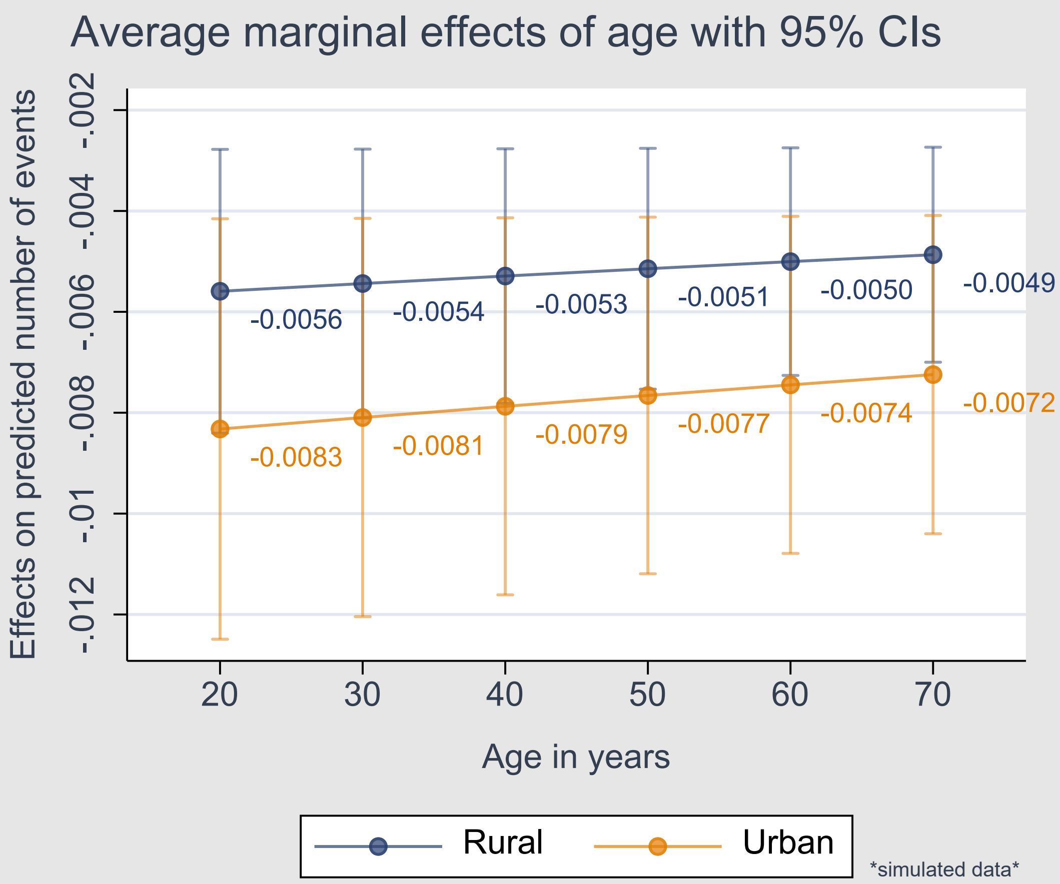
We fully break the coefficient-effect size relationship when we have a nonlinear model

- In nonlinear models, marginal effects must be taken from the derivative of the conditional expectation
  - Likewise incremental effects from the difference in conditional expectations
- Fit a Poisson model for visits instead of linear and we obtain $y = \exp(0.4705 - \mathbf{0.0028}\, age - 0.3879\, rural + 0.7205\, pvehicle)$
- Nonlinear functional forms will introduce a “shape” to the plot of marginal effects, even if the variable itself enters linearly



And, the value of other covariates now alters our effect estimates

- Fit a Poisson model for visits instead of a linear model and we obtain $y = \exp(0.4705 - \mathbf{0.0028}\, age - 0.3879\, rural + 0.7205\, pvehicle)$
- The average marginal effect of age is **−0.0072**, but
  - It differs by age, even though we don't have age^2 in our model
  - And it differs for rural vs urban, even though we have no age x residence interaction
- This will be true for any other nonlinear model you estimate
- Visualizing the relationship between a covariate and the outcome is more challenging when looking at marginal effect plots for nonlinear models — **these are changes, not values!**

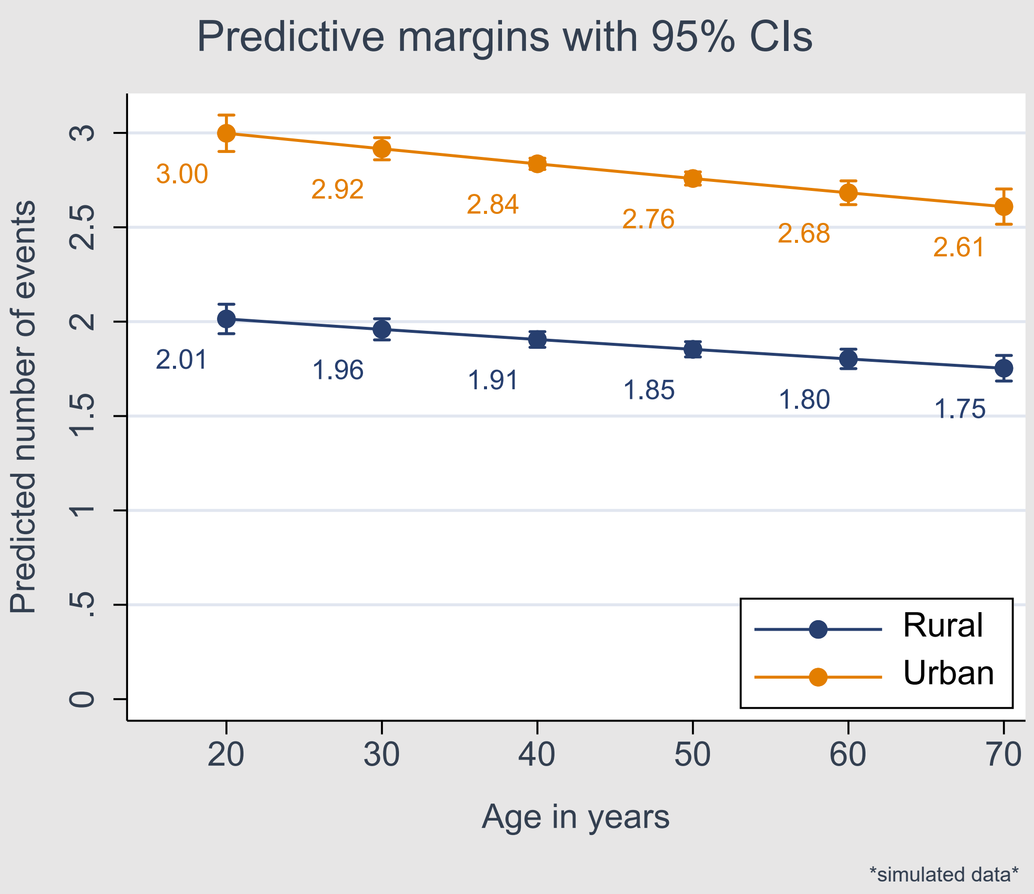


Quick review: marginal and incremental effects

- Marginal effects are calculated using the *derivative* of the conditional expectation
- Incremental effects are the *difference* in two conditional expectations
- Both marginal and incremental effects hold all other covariates “constant as observed” in calculations, but this does not mean the values of other covariates never matter
- There is one marginal (or incremental) effect estimate per variable per unique value of the variables in the derivative (or difference) equation
  - The magnitude of the estimate and the frequency of each pattern of values determines the value of any average
- Summarize information using:
  - Average marginal effect
  - Marginal effect at a representative value (including marginal effect at the mean)
- Graphs show how effect size, not the outcome, changes over values of a covariate

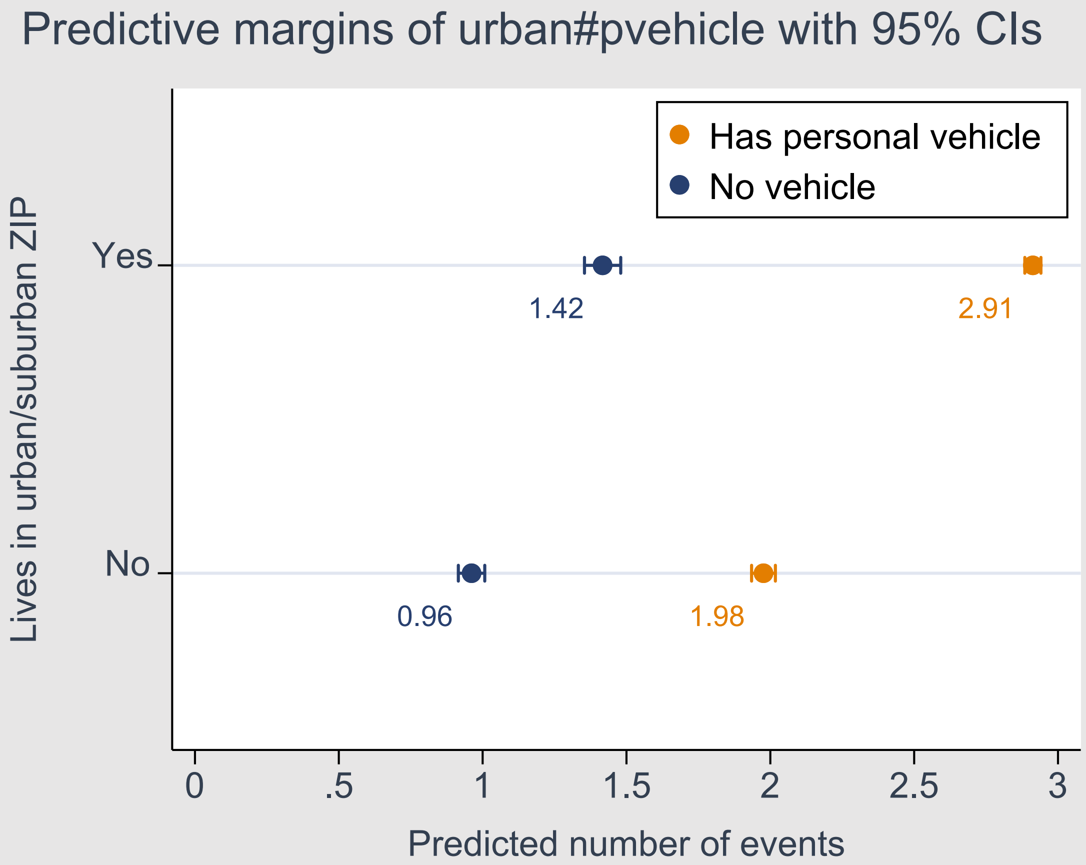
Predictive margins provide a way to visualize the relationship between the outcome and one or more covariates

- By using predictive margins, we see
  - There are significantly fewer expected visits if everyone lived in a rural area
  - The number of expected visits decreases as the population ages
  - At higher ages, the gap between expected number of visits for rural and urban areas narrows
- “Predictive margins” is a catch-all term for any conditional expectation of the outcome; variously also called
  - Marginal predictions
  - Adjusted means (continuous or semi-continuous outcomes)
  - Adjusted probabilities (discrete outcomes)



Graphing these values also helps show groupwise differences

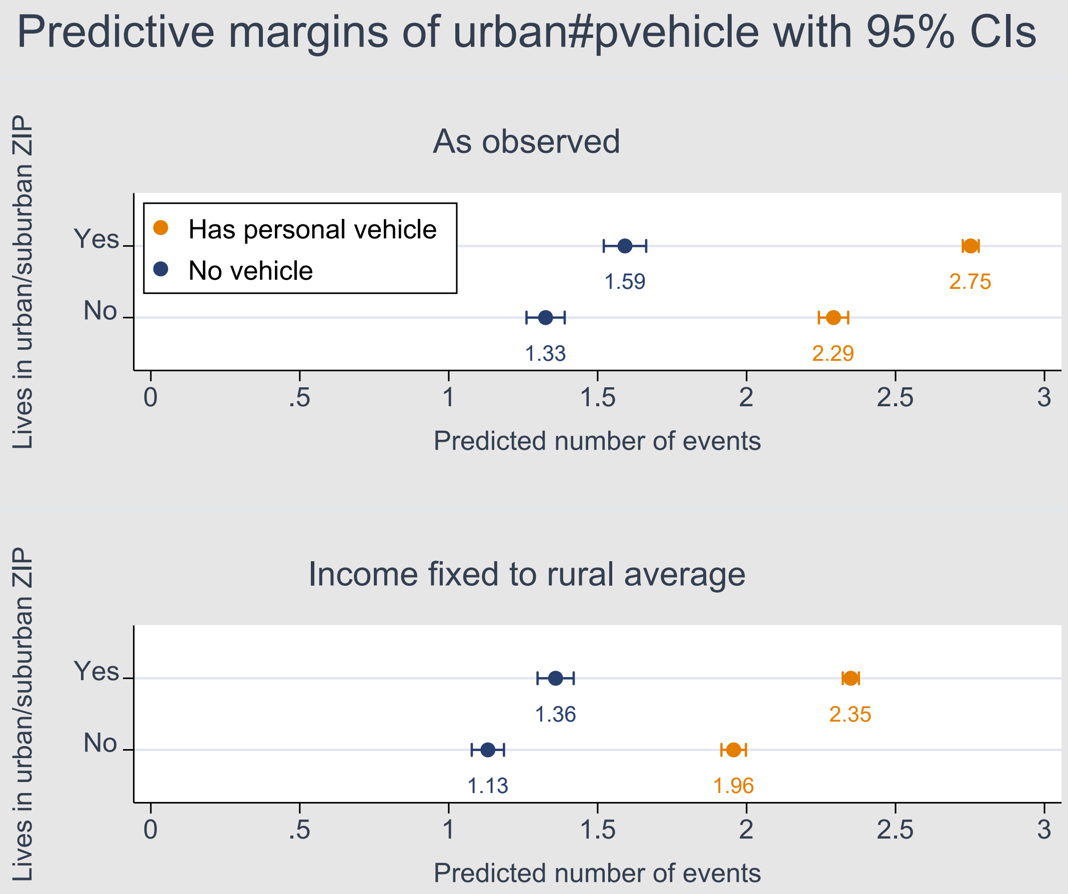
- By using predictive margins, we see
  - If everyone had a personal vehicle, the average expected number of visits would be significantly higher, regardless of location
  - Expected number of visits is lower (1.4) if everyone lives in an urban area but does not have a car versus living in a rural area with a car (1.9)
- These comparisons can be made even when there is no interaction term in the model
  - The model specifies the functional relationship (i.e., do you expect the effect to be altered by another variable?)
  - Differences in outcomes may still exist and if this is your hypothesis, look at the predictions



simulated data

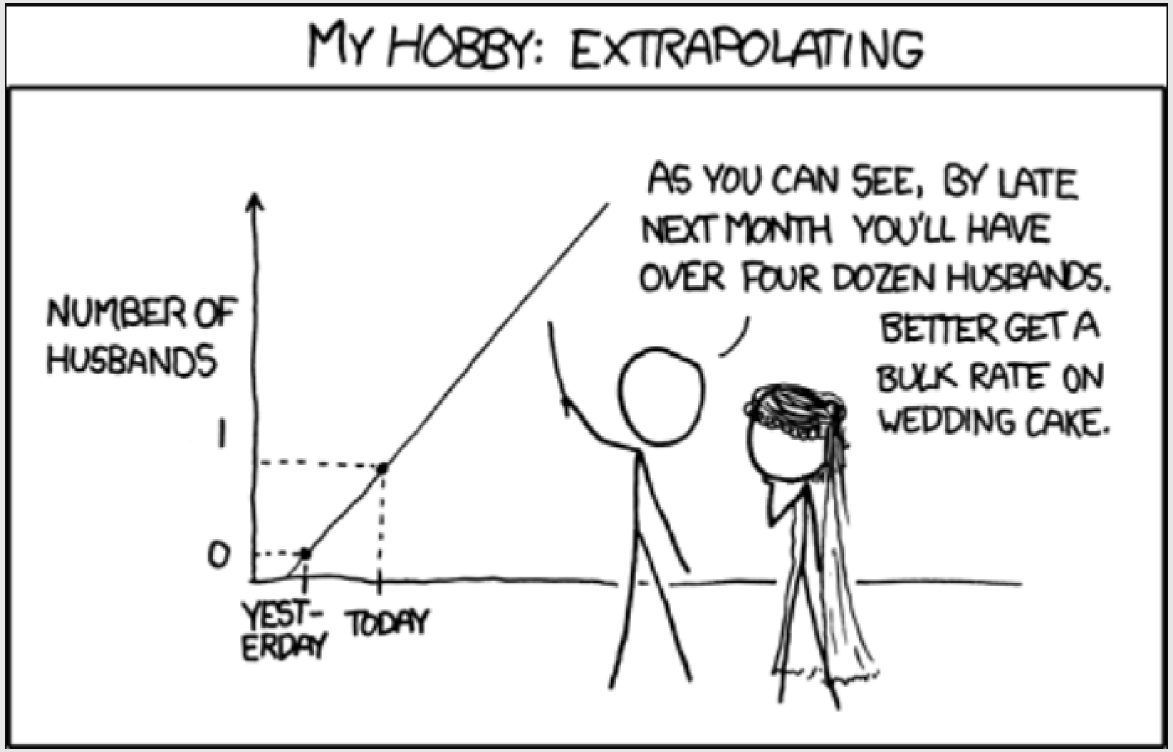
Representative values may also be used for predictive margins

- Suppose we add income to our model
  - Recall, average income in urban areas is \$162k, while it is only \$51k in rural areas
- Regardless of whether evaluating a continuous or categorical variable, taking conditional expectations means you are calculating as if everyone in the sample has the specified covariate values
  - With the default approach, predictions will be made with age and income as observed
  - With a representative value (or values), you hold group differences that are not of interest constant



simulated data

Do not extrapolate beyond the data



https://xkcd.com/605

- Whether you use marginal effects or predictive margins, keep any conditional or marginal expectation graphs constrained to the observed data range
  - You will not have reliable estimates outside this range
  - You have no guarantee that your model is appropriate outside the range
    - E.g., would we want to use a model for adults to make projections for children?

Comparing marginal effects and predictive margins

Marginal effects

- Computed from the derivative of the conditional expectation, $\partial E(y \mid z, \mathbf{w})/\partial z$
  - Where **w** is all elements of **x** except *z* & *z* is the variable we are interested in
- Also called partial effects

Predictive margins

- Computed from the conditional expectation function, $E(y \mid \mathbf{z},\mathbf{w})$
  - Where **z** is a vector of one or more covariates & **w** is all other values of **x** (as observed unless otherwise specified)
- Also called marginal predictions or adjusted predictions

Incremental effects are
$E(y \mid z=[\text{comparison}], \mathbf{w}) - E(y \mid z=[\text{base}], \mathbf{w})$

Marginal effects are best used for reporting when we are interested in the effect that a variable has on the outcome

- Always give effect sizes on the scale of the outcome of interest
  - 1 more visit on average or 4 points higher average probability of having a visit
- Easier for audiences to understand than transformations of coefficients
  - Compare "4 point increase in probability" to "1.1 times greater odds"
- Provide a convenient summary of a variable's overall effect when it appears in an interaction or the model is nonlinear
  - Average effect of age when age affects rural and urban subjects differently
- Literature documents many limitations of estimation of nonlinear transformations like the odds ratio
  - Also a great previous cyberseminar on this topic (see references)

Predictive margins are best used when our goal is to describe possible states of the world

- Always give expected average outcomes in the estimation metric
  - 2.5 visits on average or 87% average probability of having a visit
- Do not have a reference group, so expectations for all observed patterns of subject characteristics can be presented
  - Expect an average 2.7 visits if everyone lived in an urban area and only 1.7 if everyone lived in a rural area
- Can show the relationship between a variable and the outcome

More data about receipt of mental health services

- Our previous study showed that individuals with depression who live in rural areas have less use of specialty mental health care.
- We were limited by lack of data on travel barriers and facilitators.
- We have now collected more data:
  - distprov: distance in miles to the nearest mental health provider
  - transit: public transit connectivity index
- We want to know about the probability of someone getting any specialty mental health care. But, what do we want to know, <u>exactly</u>?

Our updated variables are distributed like this

	Lives in urban/suburban ZIP		
	No	Yes	Total
N (%)	4,736	15,264	20,000
	23.7%	76.3%	100.0%
Received specialty mental health care			
Yes (%)	82.6%	90.0%	88.3%
Age in years, mean (SD)	45.1	43.7	44.0
	(6.6)	(6.6)	(6.7)
Household income in $1,000s, mean (SD)	51.1	162.2	135.9
	(35.8)	(271.5)	(242.4)
Personal vehicle in household			
Yes (%)	91.0%	92.8%	92.4%
Distance to provider in miles	**20.1**	**5.0**	**8.6**
	(6.3)	**(2.2)**	**(7.4)**
Public transit connectivity index	**0.1**	**0.3**	**0.2**
	(0.2)	**(0.2)**	**(0.2)**

Poll question 3

- Which of these research questions is most interesting to you?
  - Option 1: Is the probability of receiving specialty mental health care for depression lower for residents of rural areas after controlling for other factors that could influence receiving care?
  - Option 2: What factors are associated with the largest differences in probability of receiving specialty mental health care for people with depression who live in rural versus urban areas?
  - Option 3: What is the effect of living in a rural area on the probability of receiving specialty mental health care for depression?

Choosing predictive margins is good practice when establishing that an expected outcome differs for a group

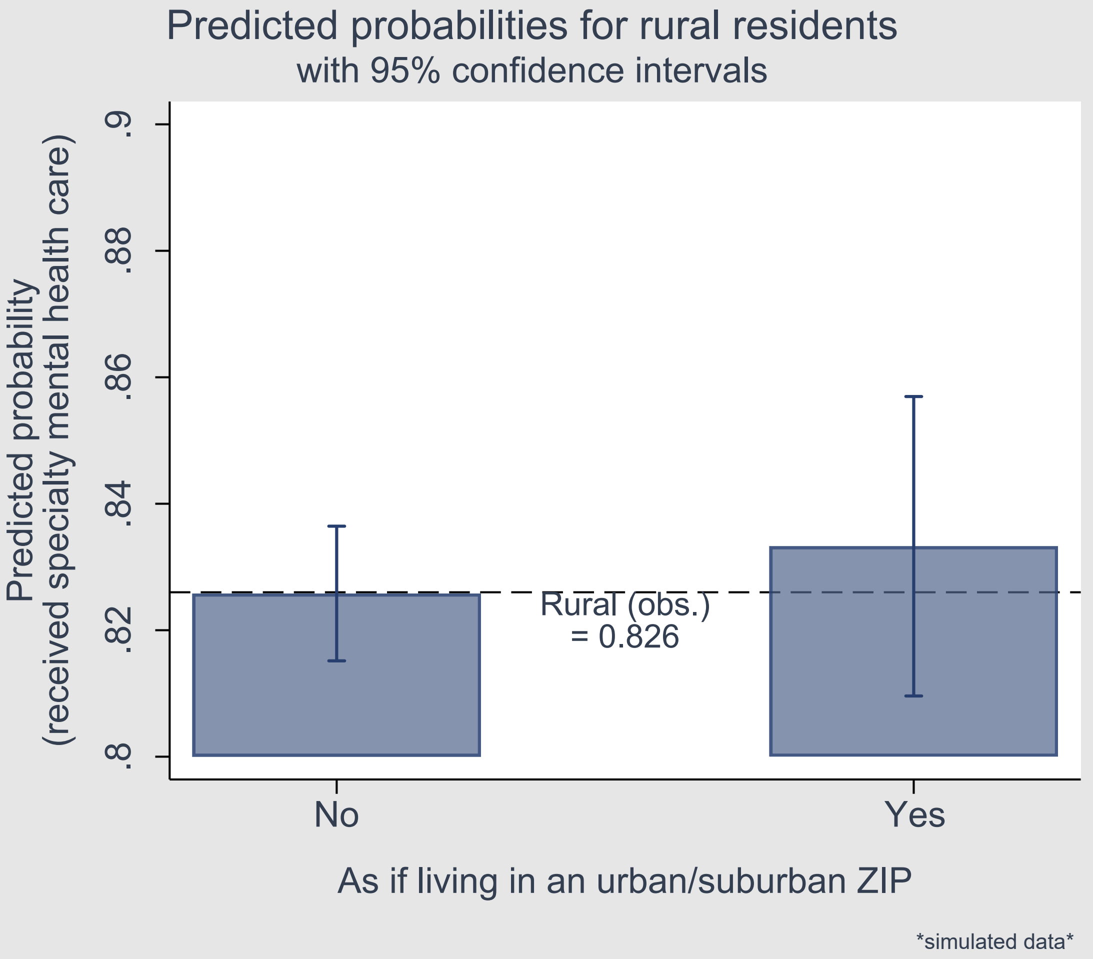
- Is the probability of receiving specialty mental health care for depression lower for residents of rural areas after controlling for other factors that could influence receiving care?
- We choose predictive margins because
  - Need to adjust for personal and area factors that differ across rural and urban areas
  - Want to determine the probability of receiving care if residents of rural areas lived in urban areas but otherwise kept their same characteristics
  - Are (possibly) not sure if we have data on all the factors related to receiving specialty mental health care

Using predictive margins to analyze expected group outcomes

- Estimate a logit model for Pr(SMHC) = *f*(rural, age, income, personal vehicle, transit connectivity, distance to provider)
- Restrict our post-estimation predictions to rural residents only
- Predict each individual's probability of receiving care if urban=0, holding all other covariates as observed, Pr(smhc=1 | urban=0, **x**), then take the average
  - Repeat for urban=1
- Because predictive margins are statistical quantities with standard errors, we can perform hypothesis tests
- Can also plot margins for a visual display

Using predictive margins to discuss expected group outcomes

- If rural residents (urban=No) lived in an urban area instead, we estimate the average probability of receiving SMHC in rural areas would be 0.833 [95% CI: 0.810 to 0.857] after adjusting for differences in age, income, personal vehicle, transit connectivity, distance to provider
  - Notice the expected probability for urban=No is equal to the observed ("unadjusted") average probability in the sample; you should always see this
  - The magnitude of the counterfactual predictive probability (here, urban=1 for the margin calculation if urban=0 in the data) will change
- Difference in predicted probabilities is not statistically significant at the 95% confidence level; Wald $\chi^2(1) = 0.32$ ($p = 0.57$)
- We conclude that moving everyone from rural to urban areas, with no other changes, would result in negligible (nonsignificant) changes in average probability of receiving care for rural residents given the factors we adjusted for

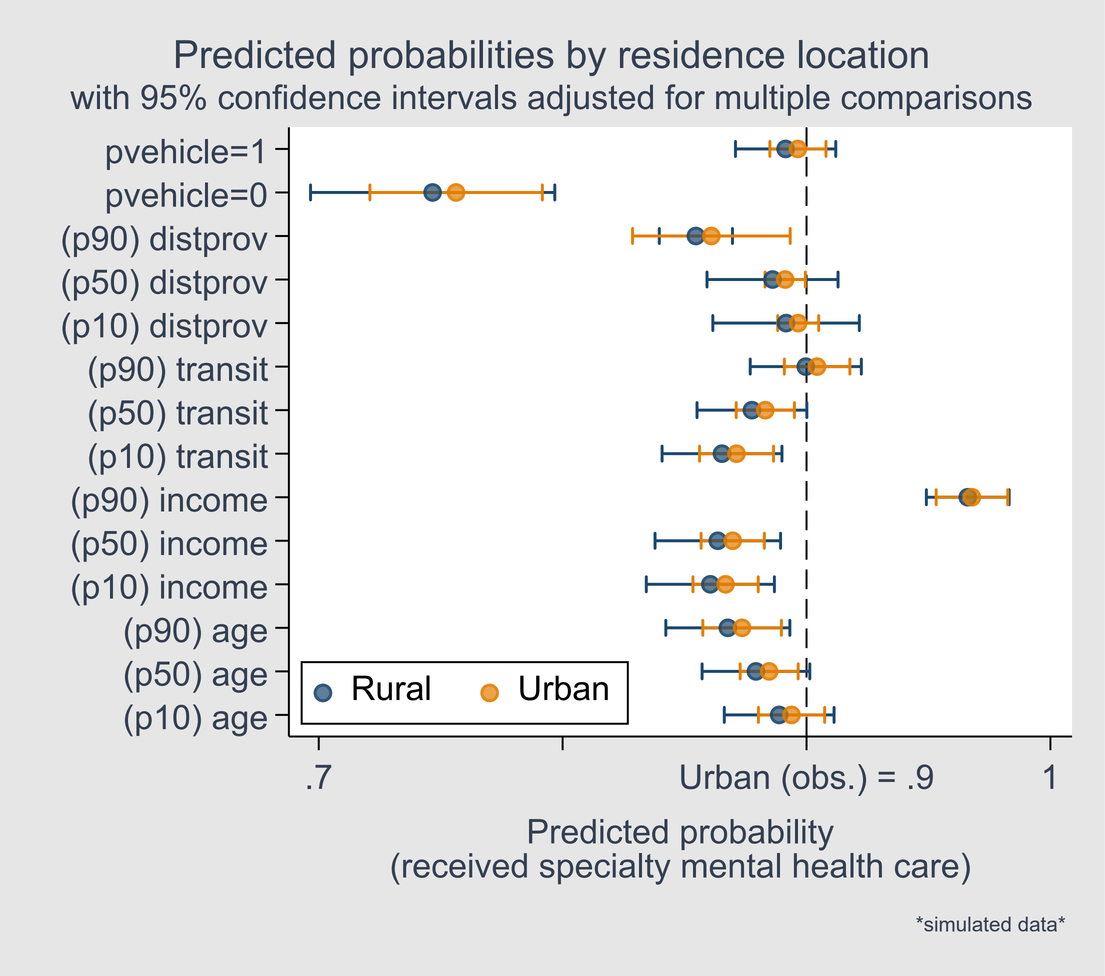


Choosing predictive margins is good practice when comparing differences in expected outcomes

- What factors are associated with the largest differences in probability of receiving specialty mental health care for people with depression who live in rural versus urban areas?
- We choose predictive margins because we
  - Want to identify potential factors that could influence receiving care and see if they are the same factors that differ by location
    - One use of analyses like this is to identify areas for future studies to target interventions
  - Are (possibly) not sure if we have data on all the factors related to receiving specialty mental health care

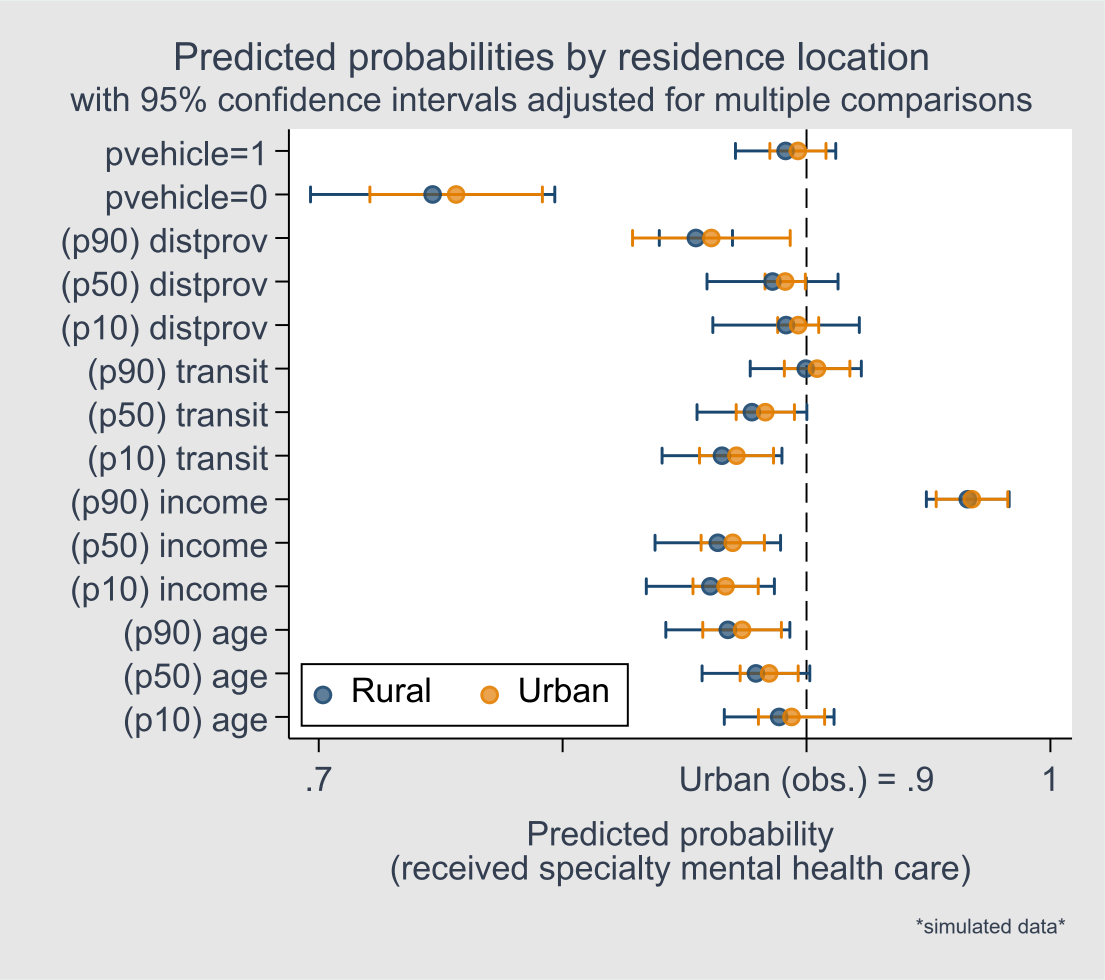
Using predictive margins to compare differences in expected outcomes

- Estimate a logit model for Pr(SMHC) = *f*(rural, age, income, personal vehicle, transit connectivity, distance to provider)
- Within levels of the other variables, compute each individual's probability of receiving care if urban=0, Pr(smhc=1 | urban=0, var=level, **w**), then take the average
  - For the only categorical variable (pvehicle) we use each possible value
  - For continuous variables, we use 10th percentile for low (e.g., "(p10) age"), median, and 90th percentile for high values
    - This is not a rule! Pick what makes sense
- Adjust confidence intervals for making multiple comparisons



Using predictive margins to discuss comparisons of differences in expected outcomes

- We previously found that rural residents tended to be slightly older, slightly less likely to have a personal vehicle, had less access to public transit, and had much lower incomes and longer travel distances on average
- Regardless of the factors examined, we do not see a significant urban-rural difference in expected probability of treatment
  - The lowest expected probability of getting treatment occurs if no one has a car
  - Other factors more common in rural areas are also associated with lower predicted probabilities
  - The highest expected probability of getting treatment occurs if everyone has income at the 90th percentile



Choosing marginal effects is good practice to obtain attributable effect sizes

- What is the effect of living in a rural area on the probability of receiving specialty mental health care for depression?
- We choose to report average marginal effects because
  - We hypothesize that something about living in a rural area leads independently to a change in probability of receiving specialty mental health care
  - We want to know what portion of the population average difference in probability of receiving care is independently attributable to living in a rural vs urban area
  - We believe that we have all covariates to model receipt of care
- If we do not think we have all covariates related to both residence and receiving care, the estimate is not interpretable as the “effect”
  - Colloquially, the “effect” language is common, but better practice to describe as a “difference” or “change” to avoid nonstatisticians linking with “cause”
  - Could also fit an alternate model, gather variables we need, etc.

Using marginal effects to obtain attributable differences

- Estimate a logit model for Pr(SMHC) = *f*(rural, age, income, personal vehicle, transit connectivity, distance to provider)
  - Use robust standard errors
- Because the urban variable is binary, calculate the incremental effect
  - Difference between predicted probability if lives in a rural area and predicted probability if lives in an urban area, with all other characteristics held as observed
- Compute the average incremental effect as the sample average of the difference
  - Use unconditional standard errors so we can make population inferences
  - Note, when the variable is a treatment indicator & all assumptions for causal inference are met, this is commonly known as the average treatment effect (ATE)

Using marginal effects to talk about attributable differences

- The average incremental effect of living in a rural area is a 0.005 decrease in probability of receiving specialty mental health care
  - If everyone lived in a rural area instead of an urban area, but otherwise kept all the same characteristics, the population average probability of receiving care would decrease by 0.005
- This effect is not statistically significant at the 95% confidence level

Marginal effects and 95% confidence limits

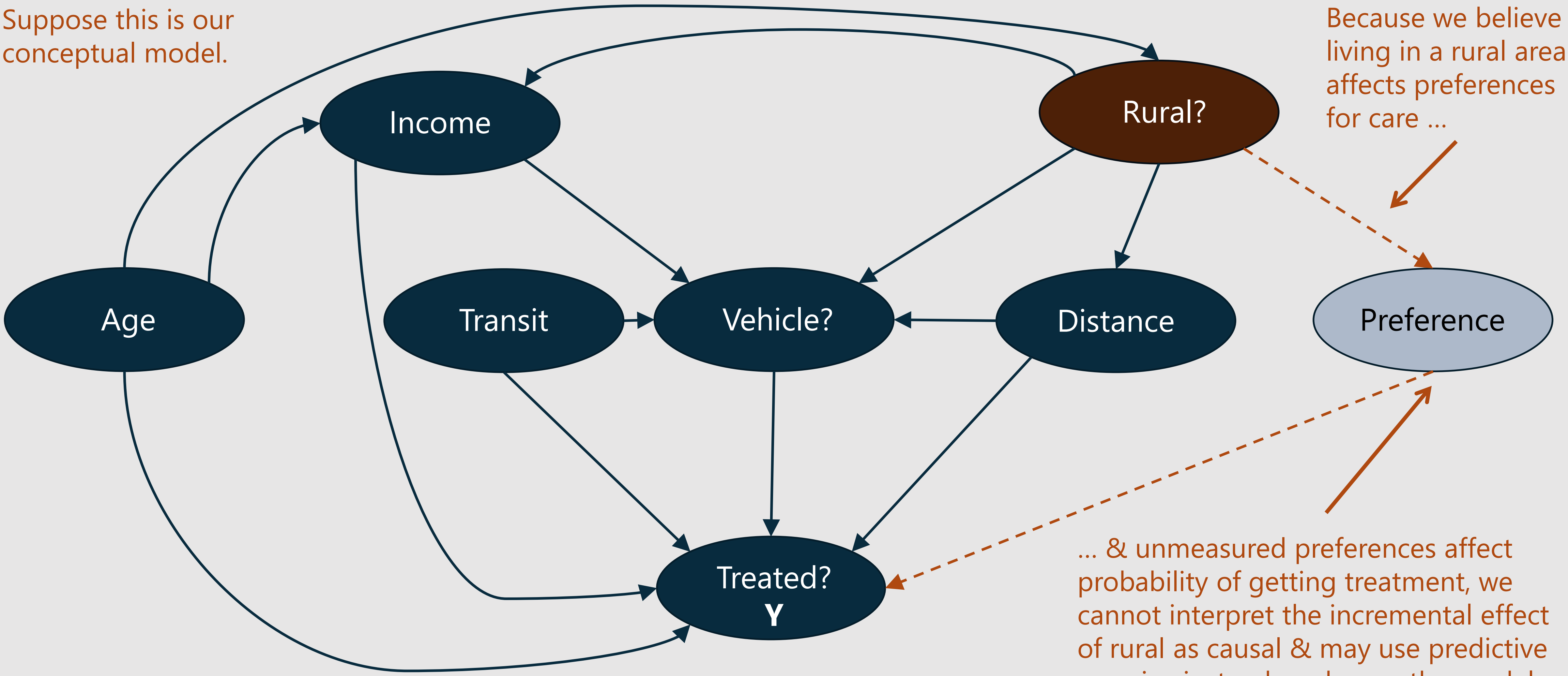
	Average marginal effects
Lives in urban/suburban ZIP	
No	**-0.005**
	[-0.025, 0.014]
Personal vehicle in household	
Yes	0.141
	[0.121, 0.162]
Age in years	-0.001
	[-0.002, -0.001]
Public transit connectivity index	0.062
	[0.041, 0.083]
Distance to provider in miles	-0.002
	[-0.003, -0.001]
Household income in $1,000s	0.001
	[0.000, 0.001]
Number of observations	20000

Incremental effects reported for categorical variables

We can use directed acyclic graphs to examine whether covariates have been omitted or may be endogenous

Suppose this is our conceptual model.

Because we believe living in a rural area affects preferences for care …



… & unmeasured preferences affect probability of getting treatment, we cannot interpret the incremental effect of rural as causal & may use predictive margins instead or change the model.

If we omit a covariate linked to rural residence, our estimate will be incorrect

- Suppose we could somehow measure preferences for seeking care
- Our point estimate for rural has changed from negative to positive
  - Not significantly different in this example dataset though
- No other variable has a major changes in its average marginal effect estimate

Average marginal effects and 95% confidence limits

	Original	No omitted variables
Lives in urban/suburban ZIP		
No	**-0.005**	**0.002**
	[-0.025, 0.014]	**[-0.017, 0.020]**
Personal vehicle in household		
Yes	0.141	0.142
	[0.121, 0.162]	[0.122, 0.162]
Age in years	-0.001	-0.001
	[-0.002, -0.001]	[-0.002, -0.000]
Public transit connectivity index	0.062	0.063
	[0.041, 0.083]	[0.042, 0.083]
Distance to provider in miles	-0.002	-0.002
	[-0.003, -0.001]	[-0.003, -0.001]
Household income in $1,000s	0.001	0.001
	[0.000, 0.001]	[0.000, 0.001]
Preference for seeking mental health care		0.048
		[0.043, 0.052]
Number of observations	20000	20000

Incremental effects reported for categorical variables

Marginal effect and predictive margin estimates are only as good as the underlying model

- Misspecification
  - Include all relevant variables that both affect the outcome and are correlated with the other variables in the model
  - Ensure the correct functional form has been selected
- Sample is not representative of the population of interest
  - Careful sample selection
  - Use sampling or post-stratification weights
  - Modeling to control for endogenous sample selection (e.g., Heckman selection model)
- Any other source of biased coefficients in the underlying model
- Investigate possible sources and interpret carefully!

Code and contact

- Annotated code and data for this talk are available for download from my GitHub trainings repository
- For questions about this talk or the code used for the talk, please email me at Rebecca.Raciborski@va.gov
- For general Stata syntax questions and advice please use Statalist (https://www.statalist.org/)
  - You can send me a private message or tag me in a public post using @Rebecca Raciborski

Other recommended related trainings

- Econometrics Seminar Series Introduction and Identification (va.gov)
- Research Design (va.gov)
- Log Odds and Ends: Marginal Effects in Logit Models (va.gov)
- Interaction Terms in Non-Linear Models (va.gov)
- Any of the HSR&D talks on advanced methods for dealing with endogeneity (e.g., instrumental variables)
- Conceptual Frameworks and Directed Acyclic Graphs (DAGs): Why We Need (and Love) Them (stanford.edu)

References and resources

- Build your own directed acyclic graph at http://dagitty.net/dags.html
- General econometrics references
  - Wooldridge, J. M. (2010). Chapter 2 of *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT press.
  - Cameron, A. C., & Trivedi, P. K. (2010). Chapter 10 of *Microeconometrics using Stata* (Vol. 2). College Station, TX: Stata press.
- Marginal effects-specific
  - Marginal Effects—Quantifying the Effect of Changes in Risk Factors in Logistic Regression Models (gentle, health-focused intro to MEs)
  - A Primer on Marginal Effects—Part II: Health Services Research Applications | SpringerLink (gives advice for different statistical packages)
    - *N.B.*: The authors reference Korn and Graubaud's survey text as a citation for predictive margins being another name for a marginal effects, but what K & G describe as a "marginal prediction" is a predictive margin; it is not a marginal effect.

Software for marginal effects (as of 4/2023)

- Stata: Use the **margins** command with the **dydx()** option after any estimation command

```
margins, dydx(*)
```

  - or **eyex()** or **eydx()** or **dxey()** if you want elasticities/semi-elasticities
  - See the [R] margins manual entry or type `help margins` in Stata

- R: First, install the `margins` package. Then, if you save parameter estimates, obtaining marginal effects is as simple as typing

```
(m <- margins(x))
```

  - R's margins is a reimplementation of Stata's `margins, dydx(*)`
  - See An Introduction to 'margins' (r-project.org)

- SAS: You may or may not be able to obtain for your model
  - Some estimation PROCs in SAS have a MARGINAL option for OUTPUT
  - Even with the MARGINAL option, you will need to compute the averages yourself and program the standard error formulas correctly
  - Consult SAS documentation for your particular PROC

Estimating incremental effects – technical appendix

- The conditional expectation for rural residence, E(y | urban=0, ***x***), is

$$\frac{\sum_{i=1}^{N} 1.4 - 0.007\, age_i + 1.88\, pvehicle_i}{N} = 1.899$$

- The conditional expectation for urban residence, E(y | rural=1, ***x***), is

$$\frac{\sum_{i=1}^{N} 1.681 - 0.007\, age_i + 1.55\, pvehicle_i}{N} = 2.799$$