Westat

Improving Lives Through Research®

# Impact Of Variable Selection on Mode Effect Adjustment for A Longitudinal Study

Xiaoshu Zhu, Ting Yan

Westat

**Federal Computer Assisted Survey Information Collection Workshops**
April 5, 2022

# Background

- Surveys using multiple modes of data collection are increasingly popular
- Concerns about potential mode effects in multimode surveys
- Mode of data collection affects resultant survey estimates through
  - Selection effect due to different respondents choosing different mode
    - E.g., in a sequential web-mail survey, younger people are more likely to choose the web mode whereas older people are more likely to choose the paper mode
  - Measurement effect due to respondents providing different answers to different mode
    - E.g., respondents are more likely to provide socially desirable answers to an interviewer-administered mode than a self-administered mode (Tourangeau and Yan, 2007)
  - An unknown mix

# Background (2)

- It is critical to understand, estimate, and adjust for mode effects in a multimode survey
- The regression modeling approach is one approach to estimate and adjust for mode effects (Kolenikov and Kennedy, 2014)
  - Regressing survey responses on mode, demographic variables, and other related variables
- This talk evaluates the impact of variable selection in the regression modeling approach on mode effect adjustment

# Methods

- Synthesized a longitudinal dataset including
  - Mode: By Phone vs. ACASI
  - Binary outcomes (past month tobacco use) measured at the current wave (time t)
    - Mode effects in these outcomes to be evaluated and adjusted
  - Variables to be included in the regression model
    - Model 1: Mode + demographic variables
    - Model 2: Model 1 + outcome measured at one prior wave of data collection (time t-1)
    - Model 3: Model 2 + one binary indicator of whether outcome changed between time t-2 and time t-1
    - Model 4: Model 2 + a count variable indicating the number of times outcome changed between time t-4 and t-1
  - Models run for adults (26+), and young adults (18-25)

# Results: Adults – Association with Mode and Outcomes

- Over 14,000 completes, Phone vs. ACASI (75% vs. 25%)
- Demographic characteristics

| | Mode | Outcomes | | | |
|---|---|---|---|---|---|
| | | Cigarette | Vaping | Cigar | Smokeless |
| **Age Category** | *** | *** | *** | * | *** |
| **Sex** | * | | | *** | *** |
| **Race/Ethnicity** | | *** | *** | *** | *** |
| **Marital Status** | ** | *** | *** | *** | *** |
| **Education Level** | *** | *** | *** | *** | *** |
| **Income Level** | *** | *** | *** | *** | |
| **Working Status** | ** | *** | *** | *** | |
| **Region** | *** | *** | * | | |

Note: * p < .05; ** p < .01; *** p < .001

# Results: Adults – Association with Mode and Outcomes (2)

› Outcome related variables

| | Mode | | | | Outcomes | | | |
|---|---|---|---|---|---|---|---|---|
| | Cigarette | Vaping | Cigar | Smoke less | Cigarette | Vaping | Cigar | Smoke less |
| **Outcome at t-1** | *** | | * | | *** | *** | *** | *** |
| **Change t-1 vs t-2** | | ** | * | | ** | *** | *** | *** |
| **Change t-1 to t-4** | | * | * | | ** | *** | *** | *** |

Note: * p < .05; ** p < .01; *** p < .001

# Results: Adults – Mode Interactions on Outcomes

- With demographic characteristics

| | Cigarette | Vaping | Cigar | Smokeless |
|---|---|---|---|---|
| **Age Category** | ** | * | | |
| **Sex** | * | | * | *** |
| **Marital Status** | * | | | |
| **Education Level** | | | * | |

Note: * p < .05; ** p < .01; *** p < .001

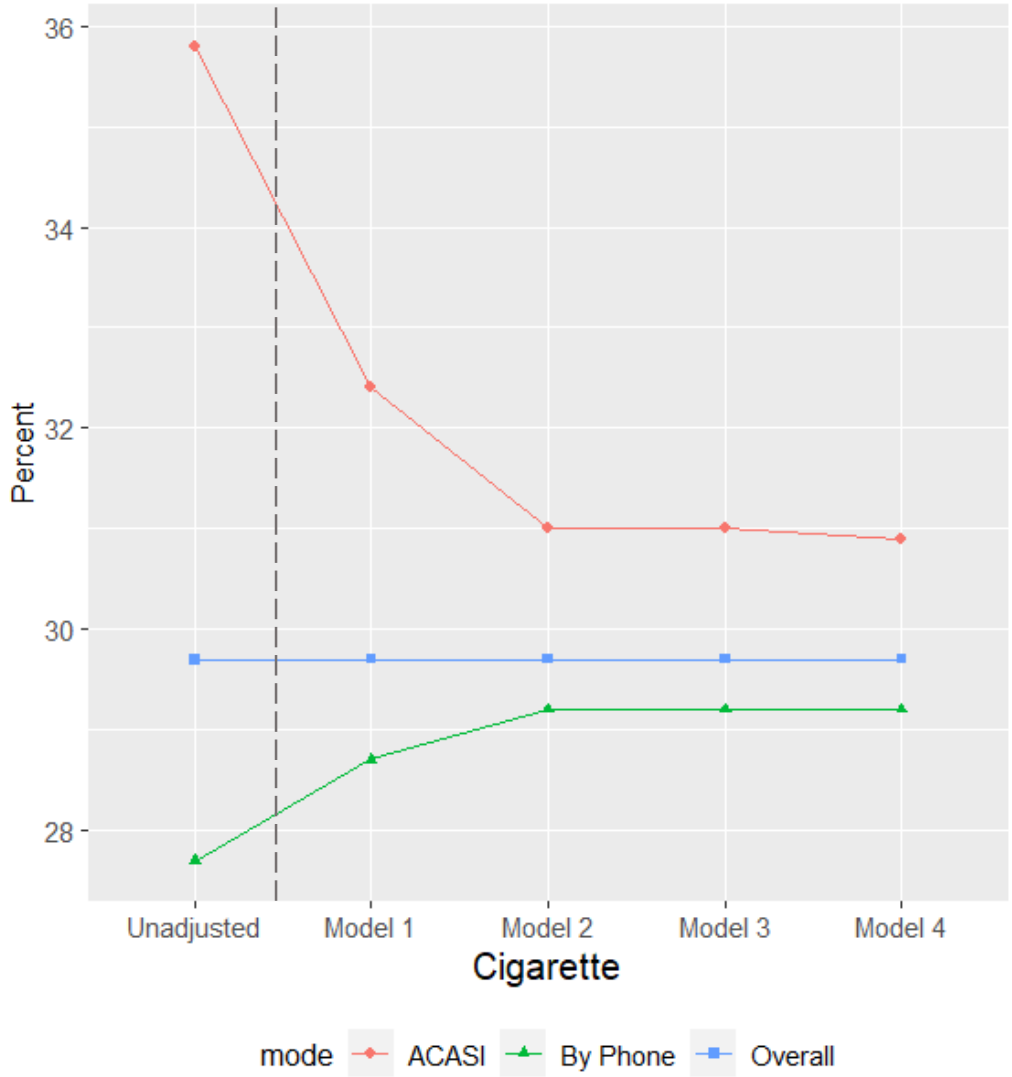
- With outcome related variables
  - No significant interactions

# Results: Adults – Model Comparison and Estimate Adjustment

## › Past Month Use: Cigarette

| | Cigarette | |
|---|---|---|
| **Unadjusted Mode Difference** | *** | |
| **Model Comparison** | **Pseudo $R^2$** | **LRT** |
| **Model 1 (Mode + Demog.)** | 0.149 | |
| **Model 2 (M1 + time t-1)** | 0.500 | *** |
| **Model 3 (M2 + t-1 vs. t-2)** | 0.501 | *** |
| **Model 4 (M2 + t-1 to t-4)** | 0.500 | *** |

Note: * p < .05; ** p < .01; *** p < .001



# Results: Adults – Model Comparison and Estimate Adjustment (2)

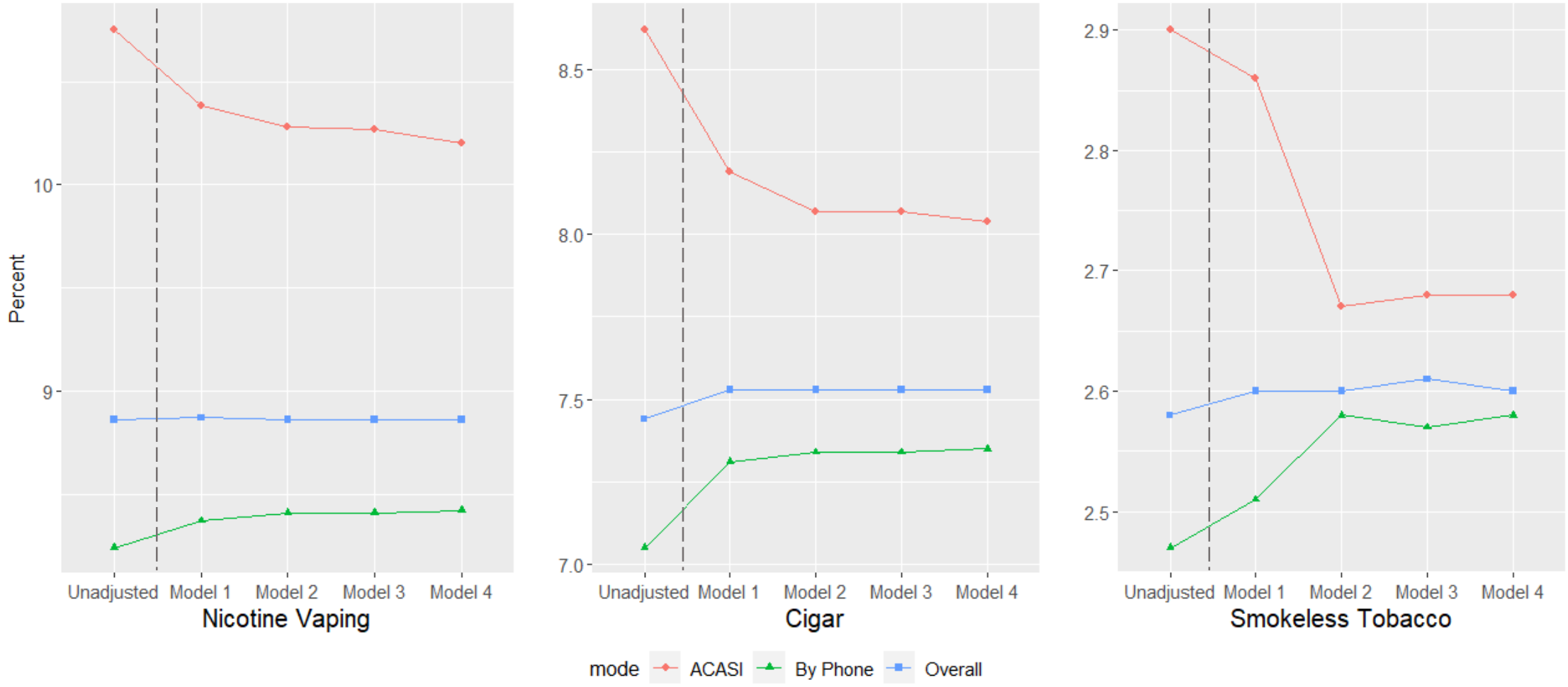
## › Past Month Use: Other tobacco products

| | Vaping | | Cigar | | Smokeless | |
|---|---|---|---|---|---|---|
| **Unadjusted Mode Difference** | *** | | ** | | | |
| **Model Comparison** | **Pseudo $R^2$** | **LRT** | **Pseudo $R^2$** | **LRT** | **Pseudo $R^2$** | **LRT** |
| **Model 1 (Mode + Demog.)** | 0.043 | | 0.025 | | 0.040 | |
| **Model 2 (M1 + time t-1)** | 0.133 | *** | 0.116 | *** | 0.132 | *** |
| **Model 3 (M2 + t-1 vs. t-2)** | 0.134 | ** | 0.116 | | 0.133 | ** |
| **Model 4 (M2 + t-1 to t-4)** | 0.137 | *** | 0.117 | *** | 0.132 | |

Note: * p < .05; ** p < .01; *** p < .001

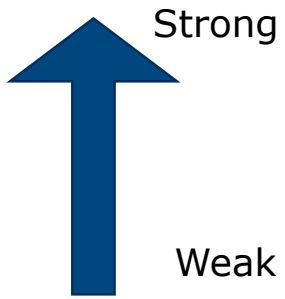
# Results: Adults – Model Comparison and Estimate Adjustment (3)

## › Estimates of Past Month Use: Other Tobacco Products



# Adults: Summary

- Pattern of mode effect and its adjustment
  - Use rate as reported in ACASI is always higher than that reported by phone
  - The gap between the outcome estimates by mode narrowed after adjustment
  - Overall estimate somehow was not affected by the variables added into the model
- Impact of variables selected to the model
  - Outcome measured at one prior wave
  - Demographic variables measured at the same wave
  - Number of changes through multiple time points
  - Change between the last two waves



# Results: Young Adults – Association with Mode and Outcomes

- Over 9,300 completes, Phone vs. ACASI (80% vs. 20%)
- Demographic characteristics

| | Mode | Outcomes | | | |
|---|---|---|---|---|---|
| | | Cigarette | Vaping | Cigar | Smokeless |
| **Age Category** | * | *** | * | * | |
| **Sex** | | * | | | ** |
| **Race/Ethnicity** | *** | ** | *** | | |
| **Marital Status** | ** | | * | | |
| **Education Level** | *** | *** | | *** | |
| **Income Level** | *** | *** | | | |
| **Working Status** | | *** | *** | ** | |
| **Region** | *** | *** | *** | ** | |

Note: * p < .05; ** p < .01; *** p < .001

# Results: Young Adults – Association with Mode and Outcomes (2)

› Outcome related variables

| | Mode | | | | Outcomes | | | |
|---|---|---|---|---|---|---|---|---|
| | Cigarette | Vaping | Cigar | Smoke less | Cigarette | Vaping | Cigar | Smoke less |
| **Outcome at t-1** | ** | | | | *** | *** | *** | *** |
| **Change t-1 vs t-2** | * | | | | ** | *** | *** | *** |
| **Change t-1 to t-4** | | | | | *** | *** | *** | *** |

Note: * p < .05; ** p < .01; *** p < .001

# Results: Young Adults – Mode Interactions on Outcomes

## Interactions with Mode

| | Past Month Use | | | |
|---|---|---|---|---|
| | Cigarette | Vaping | Cigar | Smokeless |
| **Demographic characteristics** | | | | |
| **Age Category** | | ** | | * |
| **Race/Ethnicity** | ** | | | |
| **Education Level** | *** | | | |
| **Region** | | ** | | |
| **Outcome related variables** | | | | |
| **Outcome at t-1** | | | | * |

Note: * p < .05; ** p < .01; *** p < .001

# Results: Young Adults – Model Comparison

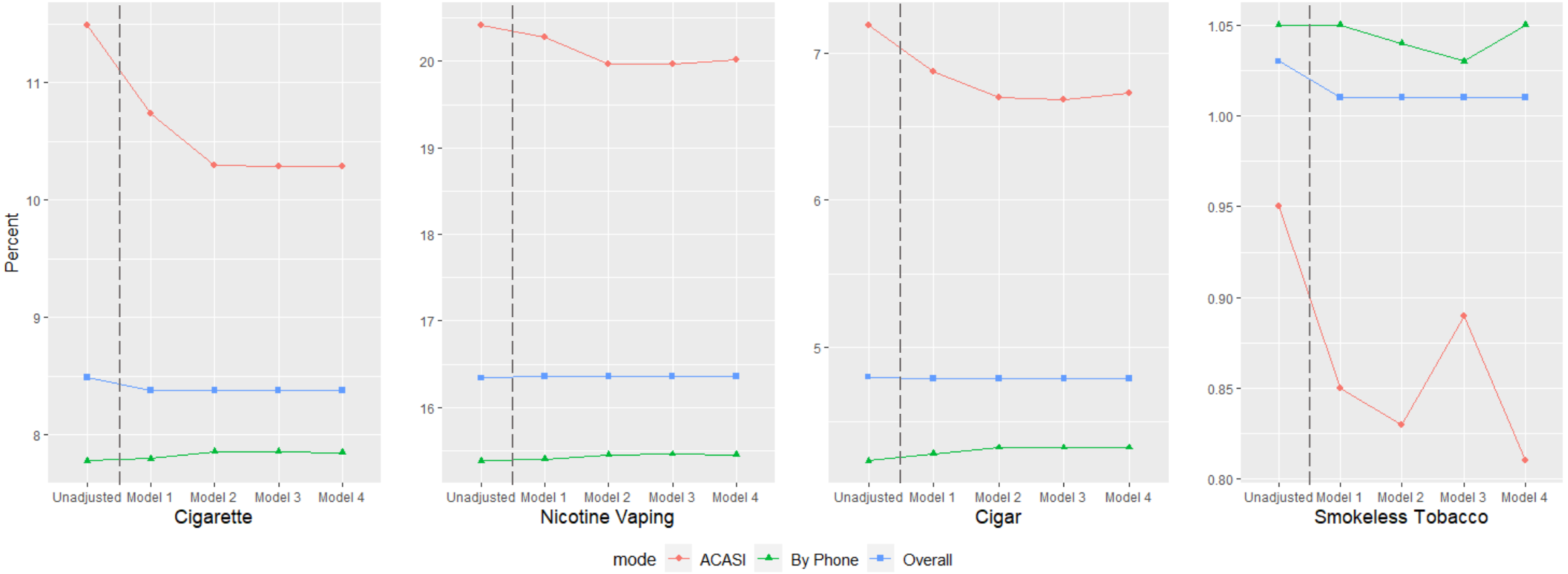
## › Past Month Use

| | Cigarette | | Vaping | | Cigar | | Smokeless | |
|---|---|---|---|---|---|---|---|---|
| **Unadjusted Mode Difference** | *** | | *** | | *** | | | |
| **Model Comparison** | **Pseudo $R^2$** | **LRT** | **Pseudo $R^2$** | **LRT** | **Pseudo $R^2$** | **LRT** | **Pseudo $R^2$** | **LRT** |
| **Model 1 (Mode + Demog.)** | 0.035 | | 0.024 | | 0.009 | | 0.005 | |
| **Model 2 (M1 + time t-1)** | 0.129 | *** | 0.152 | *** | 0.049 | *** | 0.041 | *** |
| **Model 3 (M2 + t-1 vs. t-2)** | 0.129 | | 0.152 | | 0.050 | | 0.042 | |
| **Model 4 (M2 + t-1 to t-4)** | 0.130 | ** | 0.155 | *** | 0.050 | ** | 0.041 | |

Note: * p < .05; ** p < .01; *** p < .001

# Results: Young Adults – Estimates Before and After Adjustment

› Estimates of Tobacco Product Use in Past Month



# Young Adults: Summary

- Pattern of mode effect adjustment
  - The gap between the outcome estimates by mode narrowed after adjustment but the magnitude is smaller than adults
- Impact of variables selected to the model
  - The effect of outcome measured at one prior wave is not significantly larger than the joint effect of the demographic variables measured at the same wave
  - Change between the last two waves didn’t contribute to the mode adjustment among young adults

# Conclusion

- Findings
  - Using the longitudinal information helped to narrow the gap between mode, especially the most recent information
    - Even if no mode difference on outcomes
    - Age difference on the mode effect adjustment
  - No apparent adjustment on the overall estimates
    - ACASI was affected more by adjustment, but the phone completes dominated the distribution
    - Missing variable(s) that are more informative in mode adjustment

# Conclusion (2)

- Limitation
  - Lake of randomization in the mode assignment
    - Selection effect
    - Measurement effect
  - Missing sample weight
    - hard to compare the adjusted values with the results from external surveys
- Future research
  - Search for more informative variables for mode effect adjustment
  - Other mode adjustment methods
    - Propensity matching / weighting
    - Imputation

# Reference


- Kolenikov S., and Kennedy C. (2014). Evaluating Three Approaches to Statistically Adjust for Mode Effects, *Journal of Survey Statistics and Methodology*, 2(2): 126–158. https://doi.org/10.1093/jssam/smu004
- Tourangeau, R., and Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, 133(5):859-83. doi: 10.1037/0033-2909.133.5.859.

# Thank you!

Xiaoshu Zhu
xiaoshuzhu@westat.com

Ting Yan
tingyan@westat.com