7th International Symposium on Naturalistic Driving Research

# Evaluate the Causal Relationship between Crash Risk and Cellphone Engagement Using Propensity Score Method


Danni Lu, Feng Guo

Virginia Tech Transportation Institute

Background

Confounders

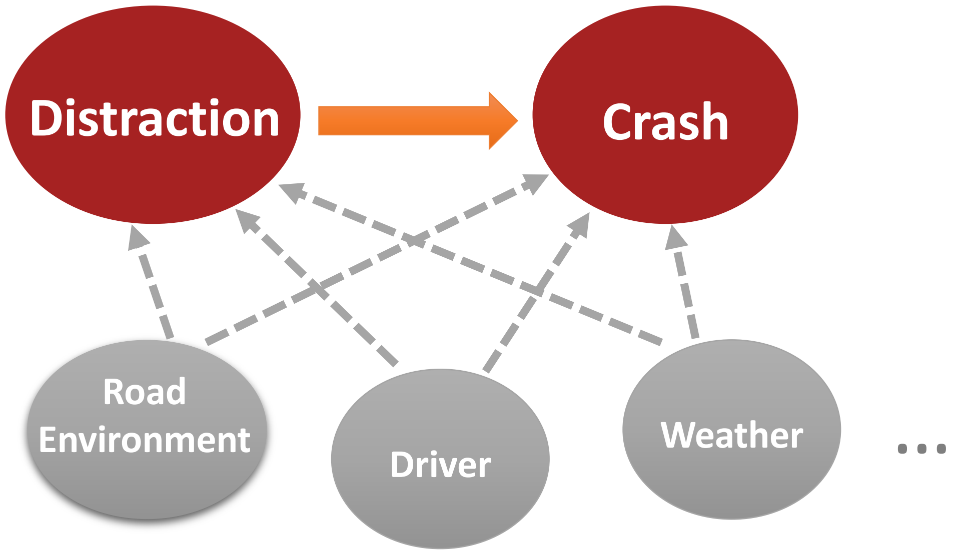
Propensity Score

Treatment Effect

Discussion

# Objectives

Evaluating the causal relationship between distraction (cellphone use) and crash risk using SHRP2 NDS data




- ✓ Evaluating confounding factors
- ✓ Exploring different propensity weighting methods

# Outline

**1** Define Response & Exposure

- Response(Y): Crash(level 1,2,3)
- Exposure(Z): cellphone use vs model driving

**2** Identify Confounding Factors(X)

- Variable selection
- Chi-square test for independence

**3** Adjust for Confounding

- Propensity score estimation
- Balance evaluation

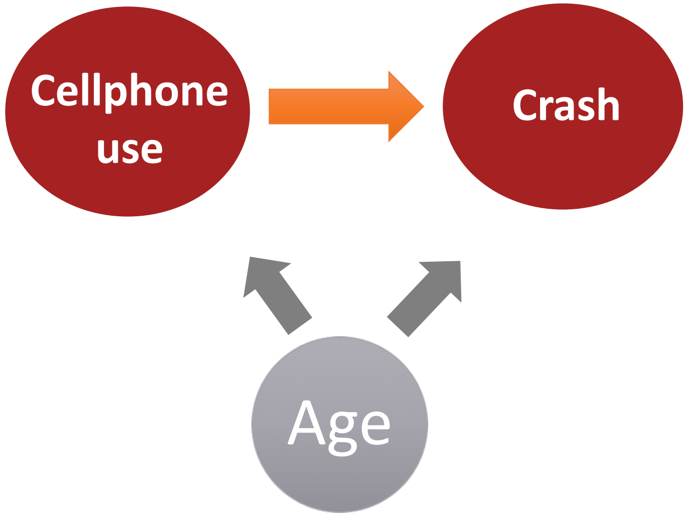
**4** Treatment Effect Estimation

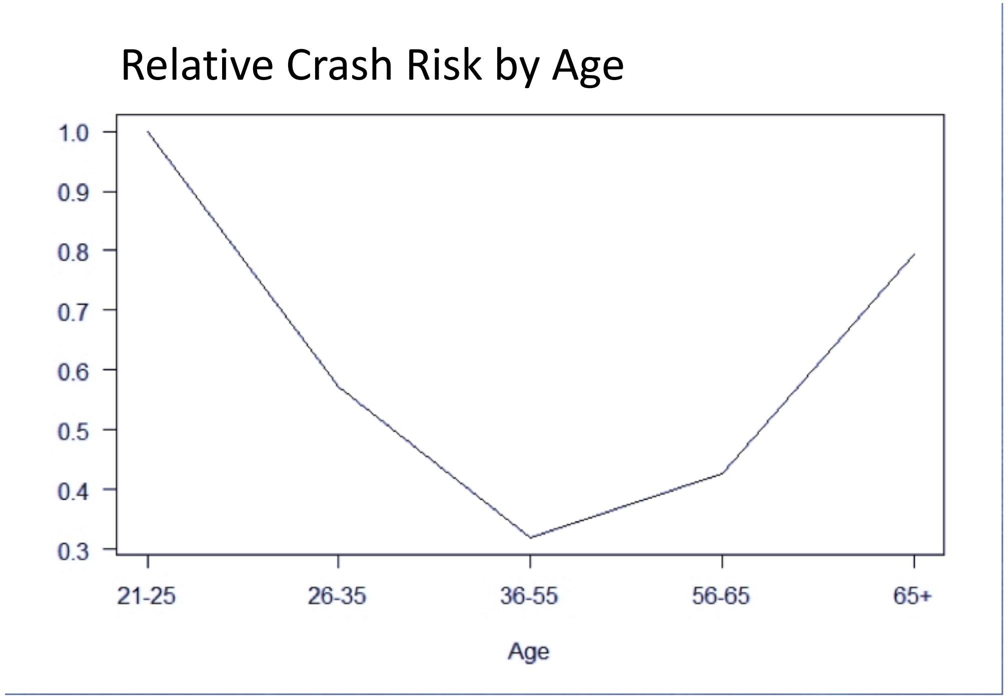
- Non-parametric estimator
- Weighted Logistic Regression
- Covariate adjustment
- Doubly Robust estimator

# Identify Confounding Factors

- Confounding

Other risk factors leading to bias in cellphone-crash relationship

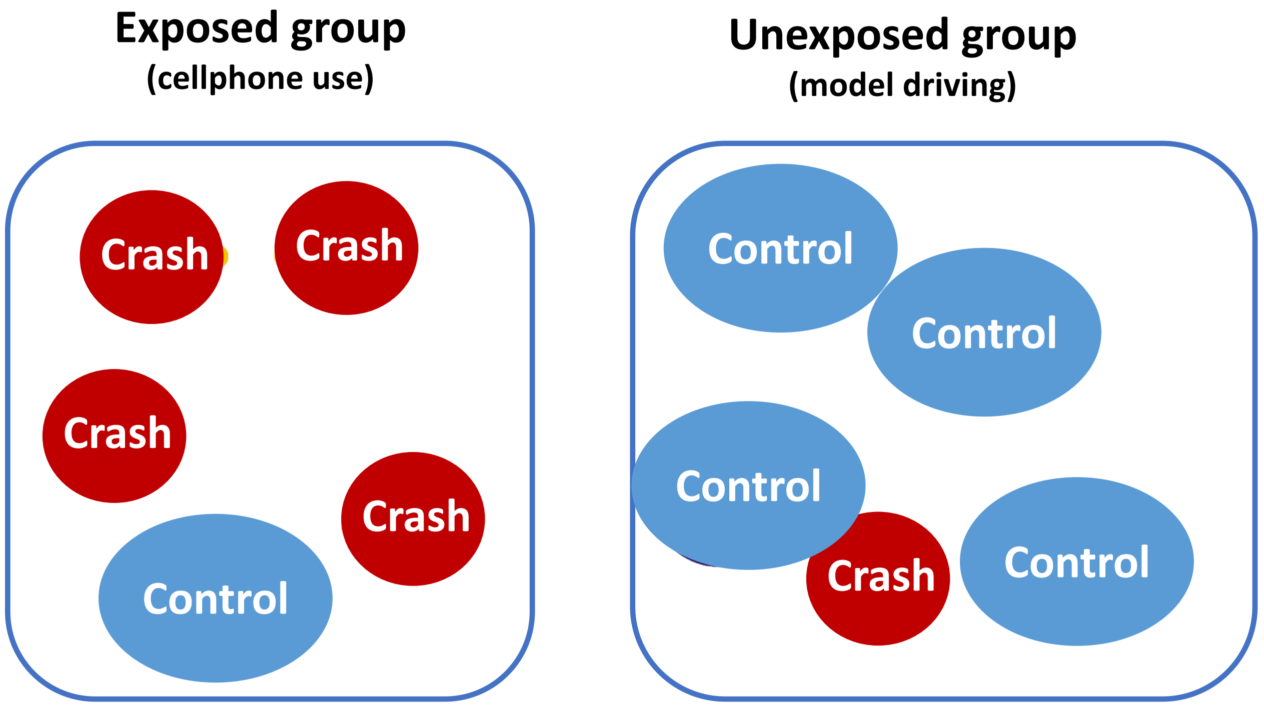


# Identify Confounding Factors

- Confounding

Other risk factors leading to bias in cellphone-crash relationship




**Propensity score:**
The probability of being assigned to treatment(exposed group) given observed confounders

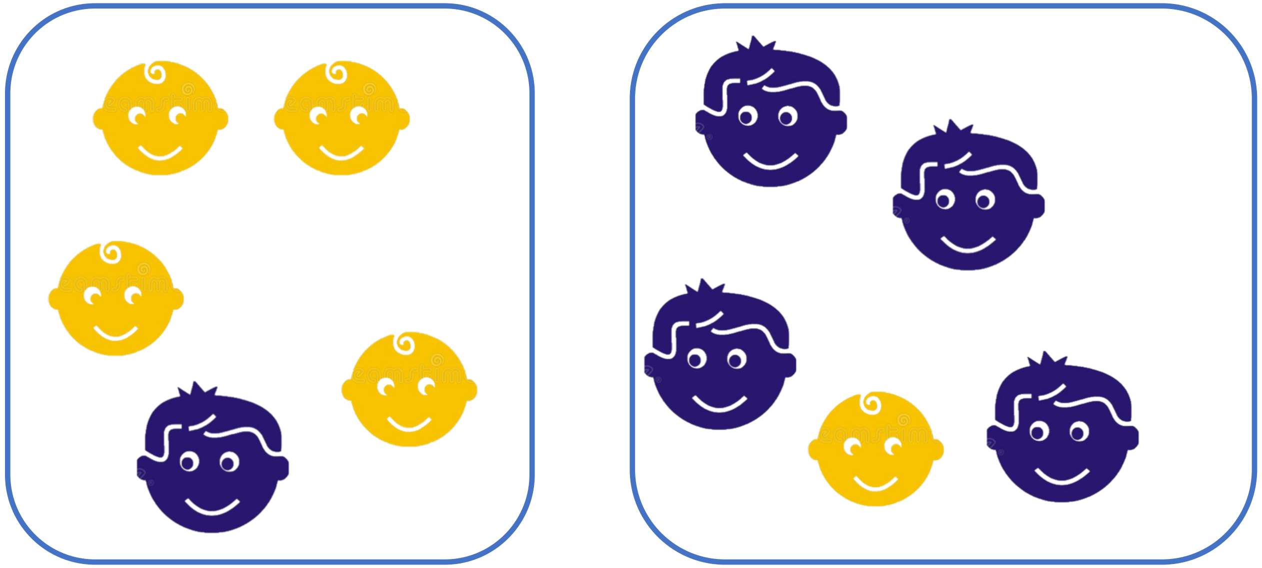
# Identify Confounding Factors

- Confounding

Other risk factors leading to bias in cellphone-crash relationship

**Exposed group**
**(cellphone use)**

**Unexposed group**
**(model driving)**



**Propensity score:**
The probability of being assigned to treatment(exposed group) given observed confounders

Background

Confounders

Propensity Score

Treatment Effect

Discussion

# Identify Confounding Factors

- Variable Selection

| Variable Type | Exposure(Cellphone) | Outcome(Crash) | Include |
|---|---|---|---|
| Confounder | correlated | correlated | ✓ |
| Instrumental variable | correlated | | ✗ |
| Related to outcome only | | correlated | ✓ |
| Related to neither | | | ✗ |

- Include variable related to outcomes to decrease the variance of estimation (Lunceford et al. 2004) ;
- Not include instrumental variables to avoid potential bias due to unmeasured confounding, as well as increase variance

# Identify Confounding Factors

- Chi-square test to identify variables correlated with outcomes and exposure

| Covariate | p value Cellphone use | p value Crash | Confounder | Instrumental Variable | Include |
|---|---|---|---|---|---|
| **Traffic density** | 0.10 | 0.00 | | | True |
| **Relation to junction** | 0.97 | 0.00 | | | True |
| **Lighting** | 0.00 | 0.00 | True | | True |
| **Age group** | 0.00 | 0.00 | True | | True |
| **Weather** | 0.08 | 0.00 | | | True |
| **Surface condition** | 0.02 | 0.00 | True | | True |
| **Traffic flow** | 0.14 | 0.00 | | | True |
| **Intersection influence** | 0.74 | 0.00 | | | True |
| **Construction zone** | 0.47 | 0.03 | | | True |
| **Income** | 0.00 | 0.04 | True | | True |
| **Locality** | 0.36 | 0.00 | | | True |
| Sex | 0.00 | 0.60 | | True | False |

Background

Confounders

Propensity Score

Treatment Effect

Discussion

# Propensity score estimation

- Propensity score

Conditional probability of being engaged with cellphone use for event *k* of driver *h*

$$e_{hk} = \Pr(Z_{hk} = 1|\boldsymbol{X}_{hk})$$

- $Z_{hk}$ : cellphone use status of event k, driver h;
- $\boldsymbol{X}_{hk}$ : observed covariates of event k, driver h;

- PS Estimation

$$logit(e_{hk}) = \delta_h + \boldsymbol{X}_{hk}\boldsymbol{\beta}$$

$\delta_h$ : random effect of driver difference

$\boldsymbol{\beta}$: regression coefficients

# Propensity score weights

- Weights:
  - Inverse probability weight(IPW)
  - Target: population average treatment effect (ATE)
  - $w_{hk} = \frac{Z}{\hat{e}_{hk}} + \frac{1-Z}{1-\hat{e}_{hk}}$

  - ATT weight
  - Target: average treatment effect on the treated(ATT)
  - $w_{hk} = Z + \frac{(1-Z)\hat{e}_{hk}}{1-\hat{e}_{hk}}$

  - ATO weight
  - Target: average treatment effect on the overlap population(ATO)
  - $w_{hk} = (1 - \hat{e}_{hk})Z + (1 - Z)\hat{e}_{hk}$

Background

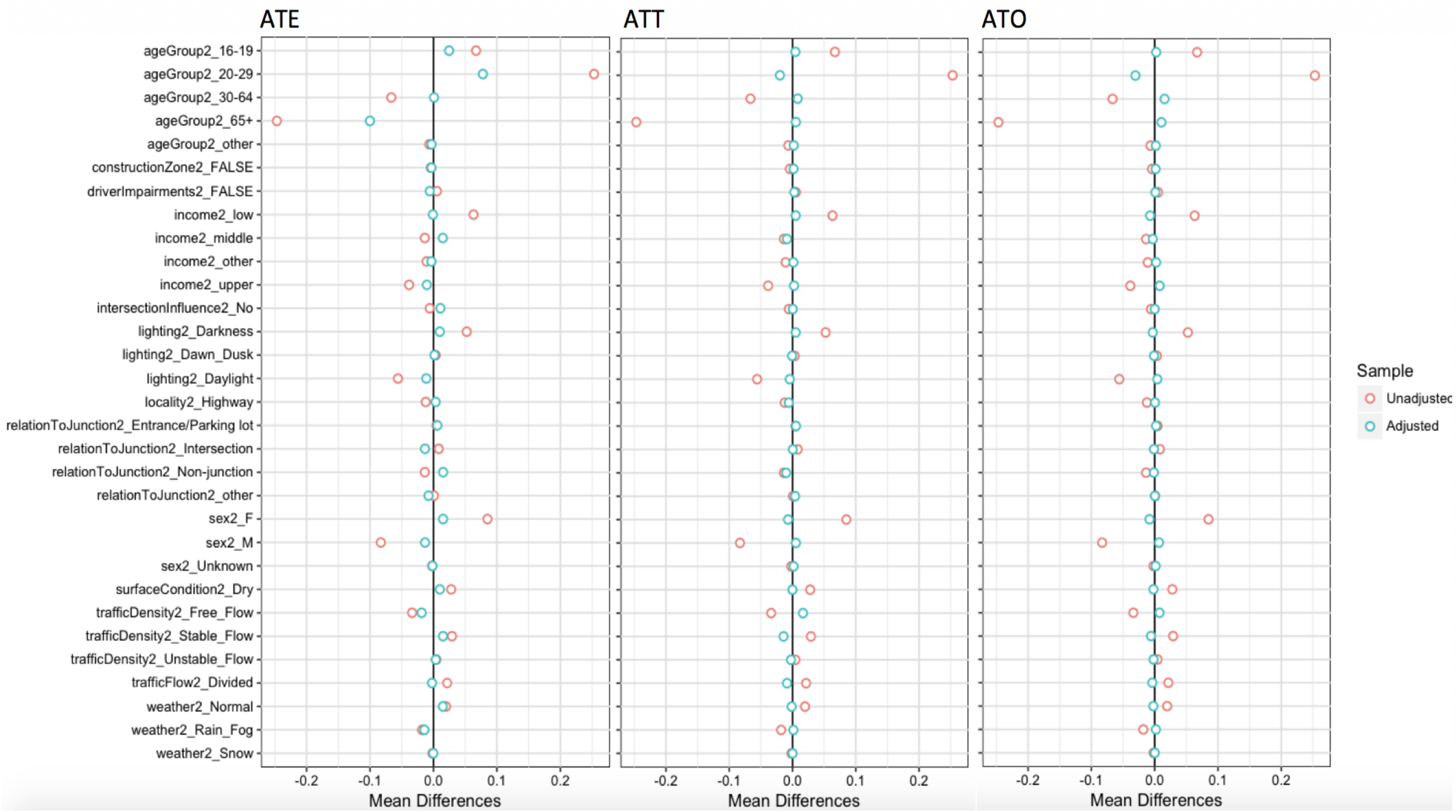
Confounders

**Propensity Score**

Treatment Effect

Discussion

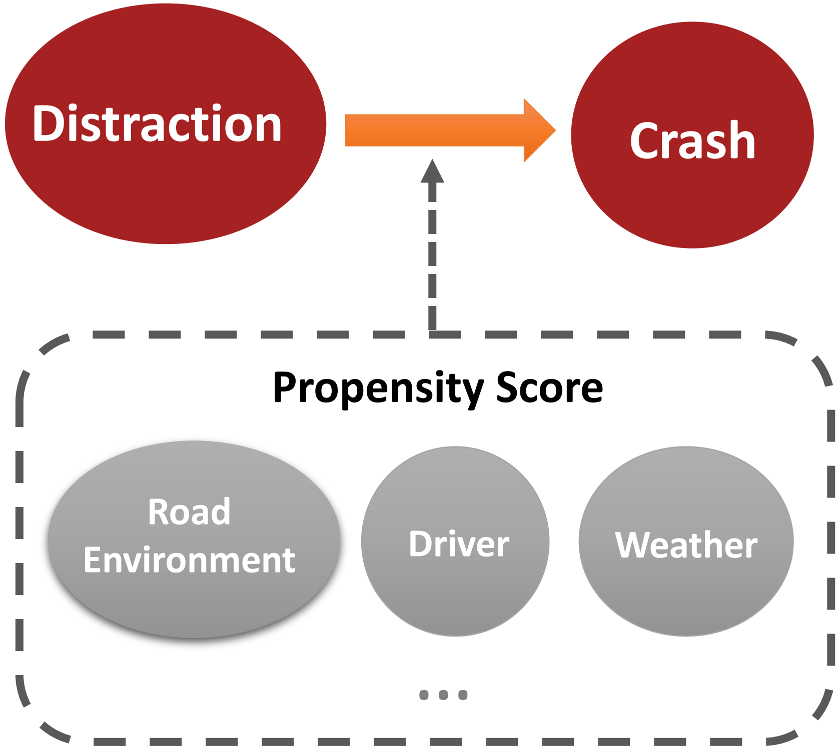
# Balance Evaluation




Standard Mean Covariate Prevalence Differences between Cellphone Use and Model Driving Groups Before/After Adjustment

# Treatment effect estimation

- What's the crash odds ratio of cellphone distraction vs model driving?




**Non-Parametric Methods:**

- Non-parametric marginal estimator
- Non-parametric clustered estimator

**Parametric Methods:**

- Weighted logistic regression
- Doubly Robust estimator
- Covariate adjustment

(For Details: see Appendix)

# Treatment effect estimation



- Methods

| | How to estimate treatment effect |
|---|---|
| • **Non-parametric marginal estimator** | Weighted average of the exposed group |
| • **Non-parametric clustered estimator** | Two steps:<br>○ Driver level weighted treatment effect;<br>○ Aggregate by driver; |
| • **Weighted GLM** | Build a weighted regression model with respect to the expected crash rate given driver and exposure status |
| • **Doubly Robust estimator** | Use parametric model to augment non-parametric estimates |
| • **Covariate adjustment** | include propensity score based weight as a additional continuous variable in the logistic regression |

# Discussion

- Challenges in estimating OR in this study:
  - Characteristics of SHRP 2 case-cohort data
    - Correlated
      - Driving behavior vary for drivers
      - Clustered estimator; random effect model
    - Rare event, Rare exposure
      - Crash is rare event, binary response variable has more zeros than ones. The prevalence of cellphone use engagement is low.
      - Doubly Robust estimator fails.

Background

Confounders

Propensity Score

Treatment Effect

Discussion

# Discussion

## ○ Methods comparison

| | Advantage | Disadvantage | SHRP2 NDS Data |
|---|---|---|---|
| **Non-parametric marginal estimator** | Easy to calculate | can not address for bias due to between cluster difference;<br>Require PS model correctly specified | Unable to exclude driver effect |
| **Non-parametric clustered estimator** | Exclude bias due to between cluster difference | Require each cluster has at least one exposed event and un-exposed event;<br>Require PS model correctly specified | Need to exclude drivers with only one exposure status |
| **Weighted GLM** | | Require PS model correctly specified | Recommended |
| **Doubly Robust estimator** | Unbiased when either outcome model or propensity score model is correct | When outcome variable is binary, can not guarantee positive estimation | Negative estimates for rare event rare exposure data |
| **Covariate adjustment** | | Can not distinguish different estimands | Target population not clear |

# Thank you!

7th International Symposium on Naturalistic Driving Research