Exchangeability Assumption in Propensity-Score Based Adjustment Methods for Population Mean Estimation Using Non-Probability Samples

Yan Li

Joint Program in Survey Methodology &
Department of Epidemiology and Biostatistics,
University of Maryland at College Park

3/1/22

Morris Hansen Lecture
This work has been generously supported by NIH-R03 CA252782

“Improving Population Representativeness of the Inference from Non-Probability Sample Analysis,” National Institute of Health

This work is an extension of two papers

Population Inference using Nonprobability Samples

- Nonprobability samples subject to Selection Bias
- Common Approaches for Improving the population representation
  - Model-based Methods
    - Regression (Wang et al. 2015)
  - Propensity Score (PB)-based adjustment
  - Doubly Robust
- Review Paper: Beaumont (2021); Rao (2021); Valliant (2020); Yang and Kim (2020)
Assumptions

- PS-based methods
  - Propensity model
  - **Conditional Exchangeability**
  - Positivity
  - Representative probability sample
  - etc…

- Model-based method
  - Outcome model
  - Transportability
  - etc…
**Notation**

- **Y**: Outcome variable of interest
- **X**: a vector of observed covariates

- **U**: the set of the finite population units of size $N$
- **C**: the set of the nonprobability sample units and $C \subset U$

**Challenge**: We observe $C$, which is NOT representing $U$

$$E_C(y) \neq E_U(y)$$
**Estimating $E(y|U)$**

- Assume Conditional Exchangeability
  
  $$E_C\{y|b(x)\} = E_U\{y|b(x)\}, \quad (\ast)$$

  where
  
  $b(x)$: a function of covariates $x$, called balancing score

- Choices of the balancing score
  
  o **Basic criteria**: Distinguish $C$ units by participation rates

  o **A natural choice**: $b(x) = P(i \in C| x, U)$

  o **Other choices**: **Finer than, if not equal to, $P(i \in C| x, U)$**
    
    - *Finest* balancing score: $b(x) = x$
    
    - *Coarest*: $b(x) = P(i \in C| x, U)$ or its monotone function
      (Rosenbaum and Rubin, 1983)
Estimation of $p(i \in C| x, U)$

- $S$: the set of a reference probability sample units with $\{x_i: i \in S\}$
- Various parametric or nonparametric models, e.g.,
  \[
  \log \left\{ \frac{p(x_i)}{1 - p(x_i)} \right\} = B^T g(x_i), \quad \text{for } i \in C \cup S, \quad (1)
  \]
  - $p(x_i)$: likelihood of being units in $C$ vs. $U$, and
    \[
    P(i \in C| x, U) = \exp(B^T g(x_i))
    \]
  - $g(x_i)$ is a known function of observed covariates
  - $B$ the unknown regression coefficients

- $\hat{B}_w$: Estimated by fitting (1) to combined $C$ and weighted $S$

- Define $b(x; \hat{B}_w) = \hat{B}_w^T g(x_i) = \log P(i \in C| x_i, U)$. Therefore,
  \[
  E_C\{y|b(x; \hat{B}_w)\} = E_U\{y|b(x; \hat{B}_w)\}
  \]
**PS-based Adjustment Estimators**

- **PS-Weighting**: Weight units in $C$ by inverse of $\exp\left(b(x, \hat{B}_w)\right)$
- **PS-Matching**: Match units in $C$ and $S$ based on $b(x; \hat{B}_w)$

**Properties**

- Approximately unbiased (Wang et al. 2020; 2021)

- **Challenge**: Variance Inflation – sample weights in $C$ vs. $S$
  (Scott and Wild, 1986)
**QUESTION**: Estimate $B$ ignoring survey weights in (1), $\hat{B}_0$,  
Define $b(x; \hat{B}_0) = \hat{B}_0^T g(x_i)$

Is $E_C\{y|b(x; \hat{B}_0)\} = E_U\{y|b(x; \hat{B}_0)\}$?

**Let us think:**

- $b(x; \hat{B}_0)$ produces sample balance in $x$ between $C$ and $S$
  
  $x \perp (C, S)|b(x; \hat{B}_0)$
  
  and therefore
  
  $E_C\{y|b(x; \hat{B}_0)\} = E_S\{y|b(x; \hat{B}_0)\}$
  
- IS $E_C\{y|b(x; \hat{B}_0)\} = E_U\{y|b(x; \hat{B}_0)\}$? Equivalently, Is $b(x; \hat{B}_0)$ a finer or monotone function of $b(x; \hat{B}_w)$? E.g. $\hat{B}_0 = \text{const.} \hat{B}_w$.
  
  GOOD LUCK!
An Adaptive Exchangeability Assumption

• 1\textsuperscript{st} step – Fit the combined sample $C \cup S$ to

$$
\log \left\{ \frac{p(i \in C)}{p(i \in S)} \right\} = \alpha + B^T g(x_i), \quad \text{for } i \in C \cup S
$$

$$
\rightarrow b(x; \hat{B}_0) = \hat{B}_0^T g(x_i)
$$

• 2\textsuperscript{nd} step – Fit the combined sample $S \cup S_w$ to

$$
\log \left\{ \frac{p(i \in S)}{p(i \in S_w)} \right\} = \gamma_0 + \gamma^T g(x_i), \quad \text{for } i \in S \cup S_w
$$

$$
\rightarrow b(x; \hat{\gamma}_w) = \hat{\gamma}_w^T g(x_i)
$$

• 3\textsuperscript{rd} step – Construct the new balancing score by adding them up

$$
b'(x) = \log \left\{ \frac{p(i \in C)}{p(i \in S_w)} \right\} = (\hat{\gamma}_w^T + \hat{B}_0^T) g(x_i), \quad \text{for } i \in C \cup S
$$
PS matching based on $b'(x)$

e.g., Kernel Weighting (KW) method by Wang et al. JRSS A 2020

$$w_{j}^{kw} = \sum_{i \in S} w_i \left( \frac{K\left(\frac{d_{ij}}{h}\right)}{\sum_{j \in C} K\left(\frac{d_{ij}}{h}\right)} \right) \quad \text{for } j \in C$$

- $w_i$ is the sample weight of survey unit $i$
- $K(\cdot)$ is an arbitrary kernel function such as standard normal
- $h$ is the bandwidth associated with $K(\cdot)$
- $d_{ij} = b'(x_i) - b_0'(x_j)$

$$\bar{y}^{kw} = \frac{\sum_{j \in C} w_j^{kw} y_j}{\sum_{j \in C} w_j^{kw}}$$
SIMULATION STUDIES

Finite population generation $U$

- N=120,000
- Three covariates $x_1, x_2, x_3 \sim N(0,1)$ with pairwise correlations $\rho_{x_1x_3} = \rho_{x_2x_3} = 0$ and $\rho_{x_1x_2} = 0.2$
- Binary outcome $Y$ with varying $\alpha_0$ with prevalence of 29%, 15% or 7%

$$P(Y = 1) = \frac{\exp(\alpha_0 + x_1\alpha_{x_1} + x_2\alpha_{x_2} + x_1x_2\alpha_{x_1x_2})}{1 + \exp(\alpha_0 + x_1\alpha_{x_1} + x_2\alpha_{x_2} + x_1x_2\alpha_{x_1x_2})}$$

Outcome predictors: $x_1$ and $x_2$
Probability Sample ($S$) & Non-probability Sample ($C$) Selection

- $n_S = 500$ and $n_C = 1500$

- Probability proportional to size sampling with measure of size
  \[
  MOS = \exp(a \times \beta^T x)
  \]

- Probability Samples with $x = (x_1, x_3)$ in MOS
  - Vary CV(weights) by setting $a = 0.1, 0.5, 1$ or $2$

- Nonprobability samples – Unknown underlying selection process
  - Quota sample on joint distributions of both $x_1$ and $x_2$
  - Quota sample on distribution of $x_1$ or $x_2$
  - Volunteer sample with unbalanced distributions in both $x_1$ and $x_2$
**PS matching estimators of population mean**

- KW with $b(x; \hat{B}_w)$ – Approx. unbiased but inflated variance
- KW with $b(x; \hat{B}_0)$ – Can be biased but more efficient
- KW with $b'(x)$ – Approx. unbiased with reduced variance

**Evaluation Criteria**

- **RelBias (%)** = (mean of 300 simulated means - population mean) divided by population mean × 100%
- **EmpVar ($\times 10^4$)** = variance of 300 simulated means
- **MSE ($\times 10^4$)** = square of bias + empirical variance
Results

1. Reference survey: (close to) self-weighted

\[ b(x; \hat{B}_0) \approx b'(x) \approx b(x; \hat{B}_w) \] due to \[ b(x; \hat{y}_w) \approx 0 \]

2. Reference survey: variable weights

a. Quota sample on joint distribution of \( x_1 \) and \( x_2 \)

\[ b(x; \hat{B}_0) \approx b'(x) \text{ more efficient than } b(x; \hat{B}_w) \]
Quota sample with balanced distribution in outcome predictors

<table>
<thead>
<tr>
<th></th>
<th>Probability samples PPS((MOS))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a = 0.1)</td>
</tr>
<tr>
<td>CV.wts</td>
<td>0.07</td>
</tr>
<tr>
<td>RelBias(%)</td>
<td></td>
</tr>
<tr>
<td>(b(x; \hat{B}_w))</td>
<td>0.34</td>
</tr>
<tr>
<td>(b'(x))</td>
<td>0.34</td>
</tr>
<tr>
<td>EmpVar</td>
<td></td>
</tr>
<tr>
<td>(b(x; \hat{B}_w))</td>
<td>1.94</td>
</tr>
<tr>
<td>(b'(x))</td>
<td>1.94</td>
</tr>
<tr>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td>(b(x; \hat{B}_w))</td>
<td>1.94</td>
</tr>
<tr>
<td>(b'(x))</td>
<td>1.95</td>
</tr>
</tbody>
</table>
b. Quota sample on a subset of predictors, $x_1$, not in $x_2$ and $x_3$

<table>
<thead>
<tr>
<th></th>
<th>Probability samples with PPS($MOS$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a = 0.1$</td>
</tr>
<tr>
<td>CV.wts</td>
<td></td>
</tr>
<tr>
<td>RelBias(%)</td>
<td></td>
</tr>
<tr>
<td>$b(x; \hat{B}_w)$</td>
<td>0.34</td>
</tr>
<tr>
<td>$b(x; \hat{B}_0)$</td>
<td>2.05</td>
</tr>
<tr>
<td>$b'(x)$</td>
<td>0.34</td>
</tr>
<tr>
<td>EmpVar</td>
<td></td>
</tr>
<tr>
<td>$b(x; \hat{B}_w)$</td>
<td>2.28</td>
</tr>
<tr>
<td>$b(x; \hat{B}_0)$</td>
<td>2.35</td>
</tr>
<tr>
<td>$b'(x)$</td>
<td>2.27</td>
</tr>
<tr>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td>$b(x; \hat{B}_w)$</td>
<td>2.29</td>
</tr>
<tr>
<td>$b(x; \hat{B}_0)$</td>
<td>2.70</td>
</tr>
<tr>
<td>$b'(x)$</td>
<td>2.27</td>
</tr>
</tbody>
</table>
Real Data Analysis

1. **COVID** with BRFSS as reference (Kalish et al. 2021)

2. **Unweighted NHANES** with NHIS as reference (Wang et al. 2021)
Data Example I – NIH SARS-CoV-2 seroprevalence study

AIM: Proportion of U.S. adults with COVID-19 antibodies from April 01 to August 04, 2020

NIH SARS-CoV-2 seroprevalence study (Kalish et al., 2021)
- More than 460,000 volunteers responding within weeks of the study announcement
- Select subset of volunteers based on age, race, sex, ethnicity and region
- A sample of 8058 subjects answered a questionnaire on medical, geographic, demographic, and socioeconomic information and provided blood samples
- Quota Sampling - Rapid data collection but suffer from Selection Bias

Behavioral Risk Factor Surveillance System (BRFSS) survey (CV(wt) = 1.92)
- A national representative probability survey
- Adjust for potential selection bias by 11 variables related to seropositivity but were not used in the quota sampling
- A total of 367,165 participants, responded to the same clinical questionnaire, were included in the analysis
<table>
<thead>
<tr>
<th>Age Group</th>
<th>Covid Survey</th>
<th>Weighted BRFSS</th>
<th>Urban/Rural</th>
<th>Covid Survey</th>
<th>Weighted BRFSS</th>
<th>Flu Vaccinated</th>
<th>Covid Survey</th>
<th>Weighted BRFSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-45</td>
<td>41.6</td>
<td>42.9</td>
<td>Urban</td>
<td>94.7</td>
<td>93.2</td>
<td>Yes</td>
<td>73.8</td>
<td>51.3</td>
</tr>
<tr>
<td>45-70</td>
<td>42.6</td>
<td>41.8</td>
<td>Rural</td>
<td>5.3</td>
<td>6.8</td>
<td>No</td>
<td>26.2</td>
<td>48.7</td>
</tr>
<tr>
<td>70-95</td>
<td>15.8</td>
<td>15.2</td>
<td>Children present</td>
<td>Yes</td>
<td>32.5</td>
<td>34.7</td>
<td>Yes</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td>67.5</td>
<td>No</td>
<td>95.9</td>
<td>90.5</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>47.4</td>
<td>47.8</td>
<td>Yes</td>
<td>32.5</td>
<td>34.7</td>
<td>Yes</td>
<td>18.8</td>
<td>18.7</td>
</tr>
<tr>
<td>Female</td>
<td>52.6</td>
<td>52.2</td>
<td>No</td>
<td>67.5</td>
<td>65.3</td>
<td>No</td>
<td>95.9</td>
<td>90.5</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White only</td>
<td>77.5</td>
<td>74.8</td>
<td>&lt;=HS</td>
<td>2.6</td>
<td>39.4</td>
<td>Yes</td>
<td>18.8</td>
<td>18.7</td>
</tr>
<tr>
<td>Black only</td>
<td>9.4</td>
<td>12.6</td>
<td>College</td>
<td>13.8</td>
<td>31.5</td>
<td>No</td>
<td>81.2</td>
<td>81.3</td>
</tr>
<tr>
<td>Others</td>
<td>13.1</td>
<td>12.5</td>
<td>&gt;=College</td>
<td>83.6</td>
<td>29.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>15.9</td>
<td>14.1</td>
<td>Own</td>
<td>75.2</td>
<td>68.8</td>
<td>Yes</td>
<td>23.4</td>
<td>31.1</td>
</tr>
<tr>
<td>Not Hispanic</td>
<td>84.1</td>
<td>85.9</td>
<td>Rent</td>
<td>20.2</td>
<td>25.6</td>
<td>No</td>
<td>76.6</td>
<td>68.9</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>16.7</td>
<td>17.1</td>
<td>Others</td>
<td>4.7</td>
<td>5.6</td>
<td>Yes</td>
<td>5.5</td>
<td>11.9</td>
</tr>
<tr>
<td>Midwest</td>
<td>15.8</td>
<td>17.6</td>
<td>Employed</td>
<td>71.2</td>
<td>57.4</td>
<td>No</td>
<td>94.5</td>
<td>88.1</td>
</tr>
<tr>
<td>Mid-Atlantic</td>
<td>20.8</td>
<td>17.3</td>
<td>NLF</td>
<td>23.8</td>
<td>32.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South/Central</td>
<td>14.2</td>
<td>15.7</td>
<td>Unemployed</td>
<td>5.0</td>
<td>10.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mountain/Southwest</td>
<td>15.5</td>
<td>15.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West/Pacific</td>
<td>17.0</td>
<td>16.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Undiagnosed seropositivity rate among US adults
04/01/2020-08/04/2020

<table>
<thead>
<tr>
<th>KW Matching</th>
<th>est (%)</th>
<th>se* (X10^-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b(x; \hat{B}_w)$</td>
<td>6.79</td>
<td>2.50</td>
</tr>
<tr>
<td>$b(x; \hat{B}_0)$</td>
<td>4.32</td>
<td>0.66</td>
</tr>
<tr>
<td>$b'(x)$</td>
<td>4.31</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Post-stratification

<table>
<thead>
<tr>
<th>KW Matching</th>
<th>est (%)</th>
<th>se* (X10^-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b(x; \hat{B}_w)$</td>
<td>4.56</td>
<td>0.83</td>
</tr>
<tr>
<td>$b(x; \hat{B}_0)$</td>
<td>4.39</td>
<td>0.61</td>
</tr>
<tr>
<td>$b'(x)$</td>
<td>4.33</td>
<td>0.61</td>
</tr>
</tbody>
</table>

*: no account for the variability due to estimating B or $\gamma$
Data Example II -- NHANES III & NHIS 1994

~ Estimate prospective 15-year all-cause mortality for people aged 18 to 75 in the US from 1990

- The Third National Health and Nutrition Examination Survey (NHANES)
  \( n_c = 17,111, \quad \widehat{N} = 173,481,294 \)
- Reference Survey: 1994 National Health Interview Survey (NHIS)
  \( n_s = 18,138, \quad \widehat{N} = 178,226,524 \) and \( \text{CV(NHIS weights)} = 0.57 \)

Both Surveys oversample old people (>= 60 yrs), minorities, low-income

Note: The two surveys share target population, data collection mode, well-designed questionnaires, and mortality information Linked to NDI.

NHIS-weighted 15-year all-cause mortality=13.04%
Estimate of 15-year Mortality Rate (%) using unweighted NHANES

<table>
<thead>
<tr>
<th></th>
<th>NHIS</th>
<th>NHANES</th>
<th>( b'(x) )</th>
<th>( b(x; \hat{B}_0) )</th>
<th>( b(x; \hat{B}_w) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sample</td>
<td>13.0</td>
<td>17.9</td>
<td>13.5%</td>
<td>16.0</td>
<td>13.4%</td>
</tr>
<tr>
<td>[18,30]</td>
<td>2.1</td>
<td>2.5</td>
<td>2.3%</td>
<td>2.3</td>
<td>2.3%</td>
</tr>
<tr>
<td>(30,50]</td>
<td>6.0</td>
<td>7.5</td>
<td>5.0%</td>
<td>5.6</td>
<td>5.0%</td>
</tr>
<tr>
<td>(50,75]</td>
<td>34.6</td>
<td>41.7</td>
<td>35.5%</td>
<td>37.8</td>
<td>35.5%</td>
</tr>
<tr>
<td>Propensity of Unweighted NHIS vs. Weighted NHIS</td>
<td>Logistic Regression of Outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>---------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimate</td>
<td>pvalue</td>
<td>Estimate</td>
<td>pvalue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age_c2</td>
<td>0.202</td>
<td>1.057</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age_c3</td>
<td>0.230</td>
<td>3.071</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>0.175</td>
<td>-0.573</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educ6</td>
<td>0.051</td>
<td>-0.065</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>race2</td>
<td>0.014</td>
<td>-0.032</td>
<td>0.597</td>
<td></td>
<td></td>
</tr>
<tr>
<td>race3</td>
<td>-0.094</td>
<td>-0.554</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>race4</td>
<td>-0.171</td>
<td>-0.855</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poverty</td>
<td>-0.123</td>
<td>-0.232</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>poverty3</td>
<td>-0.144</td>
<td>-0.064</td>
<td>0.424</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health</td>
<td>-0.020</td>
<td>0.386</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>region2</td>
<td>0.027</td>
<td>-0.040</td>
<td>0.624</td>
<td></td>
<td></td>
</tr>
<tr>
<td>region3</td>
<td>-0.059</td>
<td>0.018</td>
<td>0.801</td>
<td></td>
<td></td>
</tr>
<tr>
<td>region4</td>
<td>0.006</td>
<td>0.080</td>
<td>0.343</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marstat</td>
<td>0.450</td>
<td>0.294</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>marstat3</td>
<td>0.227</td>
<td>0.028</td>
<td>0.752</td>
<td></td>
<td></td>
</tr>
<tr>
<td>smk_stat1</td>
<td>0.003</td>
<td>0.648</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>smk_stat2</td>
<td>0.031</td>
<td>0.401</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fam_inc</td>
<td>-0.084</td>
<td>-0.164</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>snuff_chew</td>
<td>0.003</td>
<td>0.104</td>
<td>0.212</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion and Discussion

- Conditional Exchangeability (*) - balancing scores Finer than, if not equal to, the participating rate
  - Weighted propensity scores \( b(x; \widehat{B}_w) \)
  - Unweighted propensity scores \( b(x; \widehat{B}_0) \)

- Adaptive exchangeability
  - Identify \( b(x; \widehat{B}_0) \)
  - Identify bias correction factor \( b(x; \widehat{\gamma}_w) \) by comparing \( S \) vs \( S_w \),
  - Construct \( b'(x) = b(x; \widehat{B}_0) + b(x; \widehat{\gamma}_w) \),
    which a monotone function of \( P(i \in C | x, U) \).
Future Area

- Other methods to satisfy adaptive exchangeability? Poststratification?
- Variables to be collected in both C and S?
- Propensity Modeling and Estimation
  - Depends on the predictivity of propensity score model?
  - Machine learning Methods?

- High quality reference survey required by $b(x; \hat{y}_w)$
  Less variable and informative weights!

High-Quality Probability Samples are still in great demand, especially for population-level descriptive estimates
THANK YOU!
REFERENCES