Penn
UNIVERSITY of PENNSYLVANIA

Propensity Score Analysis with Survey Weighted Data

Greg Ridgeway
Department of Criminology
Department of Statistics
University of Pennsylvania

Stephanie A. Kovalchik
Victoria University
Game Intelligence Group, Tennis Australia

Beth Ann Griffin
RAND Corporation

Mohammed U. Kabeto
Department of Internal Medicine
University of Michigan

Appeared in the *Journal of Causal Inference*, 2015

Use Survey Weights at All Stages

Use Survey Weights at All Stages

- use the sampling weights in the propensity score model
- use the sampling weight times the propensity score weight in the final outcome analysis

Outline

- Background
- Derivation
- Risks that proper use of sampling weights avoid
- Simulation and analysis of police survey data
- Conclusions

Outline

- Background
- Derivation
- Risks that proper use of sampling weights avoid
- Simulation and analysis of police survey data
- Conclusions

Applied Researchers Confused About Propensity Scores and Survey Weights

- DuGoff, Schuler, and Stuart (2014) found 28 health services research studies with design weights and analyses involving propensity scores
    - 16 ignored the weights completely
    - 7 used the weights only in the outcome model
    - 5 used the weights in both the propensity score and outcome model

Statistical Literature Offers Conflicting Advice

> "we argue that the propensity score model **does not need to be survey-weighted**, as we are not interested in generalizing the propensity score model to the population"

Statistical Literature Offers Conflicting Advice

correct analyses depend on analysts considering "the joint distribution of the observations and of the sampling and assignment indicator variables"

"recommend including the sampling weight as a predictor in the propensity score model"

Outline

- Background
- Derivation
- Risks that proper use of sampling weights avoid
- Simulation and analysis of police survey data
- Conclusions

Average Treatment Effect on the Treated Is Today's Focus

$$\hat{E}(y_1 - y_0 | t = 1) = \frac{\sum_{i=1}^{n} t_i \frac{1}{p_i} y_{1i}}{\sum_{i=1}^{n} t_i \frac{1}{p_i}} - \frac{\sum_{i=1}^{n} (1 - t_i) w_i y_{0i}}{\sum_{i=1}^{n} (1 - t_i) w_i}$$

- t_i is a 0/1 treatment indicator
- y_{1i} is the treatment outcome of case *i*
- y_{0i} is the control outcome of case *i*
- p_i is the sampling probability
- w_i combines sampling weights and propensity score weights

Weights That Align $f(\mathbf{x})$ for Treated and Sampled Controls

$$f(\mathbf{x}|t=1) = w(\mathbf{x}) f(\mathbf{x}|t=0, s=1)$$

- Rearranging + Bayes Theorem

$$w(\mathbf{x}) = \frac{f(s=1|t=0)}{f(t=0)} \frac{1}{f(s=1|t=0,\mathbf{x})} \frac{f(t=1|\mathbf{x})}{1-f(t=1|\mathbf{x})}$$

- $f(t=1|\mathbf{x}) \neq f(t=1|\mathbf{x}, s=1)$

$$\hat{E}(y_0|t=1) = \frac{\sum_{i=1}^{n}(1-t_i)\frac{1}{p_i}\frac{e_i}{1-e_i} y_{0i}}{\sum_{i=1}^{n}(1-t_i)\frac{1}{p_i}\frac{e_i}{1-e_i}}$$

Outline

- Background
- Derivation
- Risks that proper use of sampling weights avoid
- Simulation and analysis of police survey data
- Conclusions

Risk #1: Missing Sampling Weight Variable in Propensity Score

- Sampling and treatment assignment depend on *z*, but potential outcomes do not depend on *z*

N	***x***	***z***	$P(s=1\|x,z)$	$P(t=1\|x,z)$	$E(y_0\|x,z)$	$E(y_1\|x,z)$
1,000	0	0	0.2	0.1		
1,000	0	1	0.3	0.9		
1,000	1	0	0.4	0.8	4	4
1,000	1	1	0.5	0.8	4	4

$P(t=1|x=0)=0.50$

$P(t=1|x=0,s=1)=0.58$

	$t=1$	$t=0$	$t=0$
Sampling weight PS model			
Sampling weight outcomes	Yes		
$\mathrm{E}(x\|t)$	0.615		
$\mathrm{E}(y_t\|t=1)$	2.846		

Risk #1: Missing Sampling Weight Variable in Propensity Score

- Sampling and treatment assignment depend on *z*, but potential outcomes do not depend on *z*

N	***x***	***z***	$P(s=1\|x,z)$	$P(t=1\|x,z)$	$E(y_0\|x,z)$	$E(y_1\|x,z)$
1,000	0	0	0.2	0.1	1	1
1,000	0	1	0.3	0.9	1	1
1,000	1	0	0.4	0.8	4	4
1,000	1	1	0.5	0.8	4	4

	$t=1$	$t=0$	$t=0$
Sampling weight PS model		Yes	
Sampling weight outcomes	Yes	Yes	
$\mathrm{E}(x\|t)$	0.615	0.615	
$\mathrm{E}(y_t\|t=1)$	2.846	2.846	

Risk #1: Missing Sampling Weight Variable in Propensity Score

- Sampling and treatment assignment depend on *z*, but potential outcomes do not depend on *z*

N	***x***	***z***	$P(s=1\|x,z)$	$P(t=1\|x,z)$	$E(y_0\|x,z)$	$E(y_1\|x,z)$
1,000	0	0	0.2	0.1	1	1
1,000	0	1	0.3	0.9	1	1
1,000	1	0	0.4	0.8	4	4
1,000	1	1	0.5	0.8	4	4

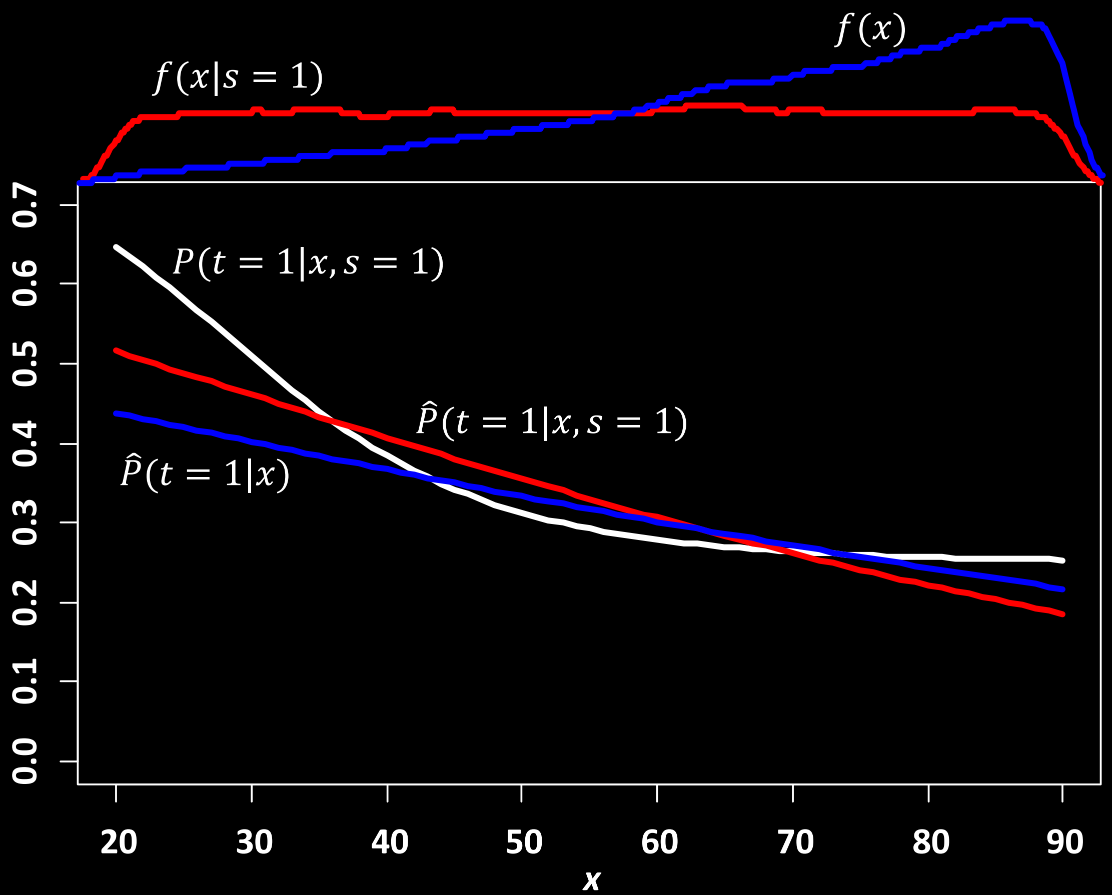
	$t=1$	$t=0$	$t=0$
Sampling weight PS model		Yes	No
Sampling weight outcomes	Yes	Yes	Yes
$\mathrm{E}(x\|t)$	0.615	0.615	0.537
$\mathrm{E}(y_t\|t=1)$	2.846	2.846	2.610

Risk #2: Spending Degrees of Freedom in the Wrong Places

- Recommend using modern statistical methods for estimating propensity scores, such as gbm/fastDR
- Quality of the propensity score will matter most for **x** with large sampling weight

$$w(\mathbf{x}) = \frac{1}{f(s=1|t=0,\mathbf{x})} \frac{f(t=1|\mathbf{x})}{1-f(t=1|\mathbf{x})}$$

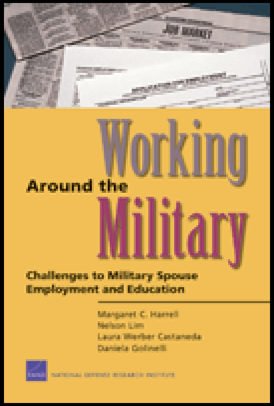
Risk #2: Spending Degrees of Freedom in the Wrong Places

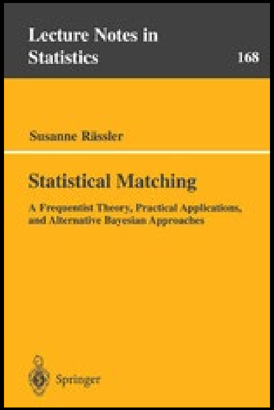


- Treatment group mean = 65.45
- Control mean w/o sampling weights = 63.48
- Control mean with sampling weights = 65.67

Risk #3: Weighted Samples Drawn from Different Sources

- Data fusion matches a collection of cases in two data sources that have similar features
  - Harrell et al (2004) compared military spouses with similar members of the general public
  - Rässler (2002) compared television viewing and consumer behavior
- Respondents with the same weight will *not* share the same features
- Sampling probability depends on the treatment assignment





Outline

- Background
- Derivation
- Risks that proper use of sampling weights avoid
- Simulation and analysis of police survey data
- Conclusions

Simulation Varied Relationships Between x, t, s, y_0, and y_1

- DuGoff, Schuler, and Stuart (2014) simulation
  - $X_i \sim N(\mu_j, 1)$
    - 30,000 from $\mu_1 = -\frac{1}{4}$
    - 30,000 from $\mu_2 = 0$
    - 30,000 from $\mu_3 = \frac{1}{4}$
  - $\text{logit}\, P(t = 1|x) = -1 + 1.39x$
  - $\text{logit}\, P(s = 1|x, t) = -2.8 - 1.39x$
  - $y_0 \sim N\left(1 + x, \frac{1}{4}\right)$ and $y_1 \sim N\left(y_0 + 0.2 + 0.1x, \frac{1}{4}\right)$

Considered Alternate Scenarios of Nonlinearity and Dependence

- DuGoff, Schuler, and Stuart (2014) simulation
  - $X_i \sim N(\mu_j, 1)$
    - 30,000 from $\mu_1 = -\frac{1}{4}$
    - 30,000 from $\mu_2 = 0$
    - 30,000 from $\mu_3 = \frac{1}{4}$
  - $\text{logit}\, P(t = 1|x) = -1 + 1.39x^2$
  - $\text{logit}\, P(s = 1|x, t) = -2.8 - 0.69x - 0.69t$
  - $y_0 \sim N\left(1 + x, \frac{1}{4}\right)$ and $y_1 \sim N\left(y_0 + 0.2 + 0.1x, \frac{1}{4}\right)$

Considered Alternate Scenarios of Dependence and Nonlinearity

	Standard	Random sample	Selection depends on (x, t)	Weight scales differ	Non-linear treatment
Select	$s \sim x$	$s \perp x$	$s \sim (x, t)$	$s \sim x \vert t$	$s \sim x$
Treatment	$t \sim x$	$t \sim x$	$t \sim x$	$t \sim x$	$t \sim x^2$

Considered Four Different Approaches to Estimation

		Standard	Random sample	Selection depends on (x, t)	Weight scales differ	Non-linear treatment
	Select	$s \sim x$	$s \perp x$	$s \sim (x, t)$	$s \sim x \mid t$	$s \sim x$
	Treatment	$t \sim x$	$t \sim x$	$t \sim x$	$t \sim x$	$t \sim x^2$
Sampling weights	PS model					
No	$t \sim 1$					
No	$t \sim x$					
Covariate	$t \sim x + sw$					
Weight	$t \sim x$					

Using Sampling Weights Consistently Provides Good Covariate Balance, Standardized Mean Difference in *x*

		Standard	Random sample	Selection depends on (*x*, *t*)	Weight scales differ	Non-linear treatment
	Select	$s \sim x$	$s \perp x$	$s \sim (x, t)$	$s \sim x \mid t$	$s \sim x$
	Treatment	$t \sim x$	$t \sim x$	$t \sim x$	$t \sim x$	$t \sim x^2$
Sampling weights	PS model					
No	$t \sim 1$	1.18	1.06	1.10	1.17	0.13
No	$t \sim x$	0.17	0.03	0.08	0.16	0.74
Covariate	$t \sim x + sw$	0.12	0.03	1.57	0.44	1.10
Weight	$t \sim x$	0.12	0.02	0.06	0.11	0.08

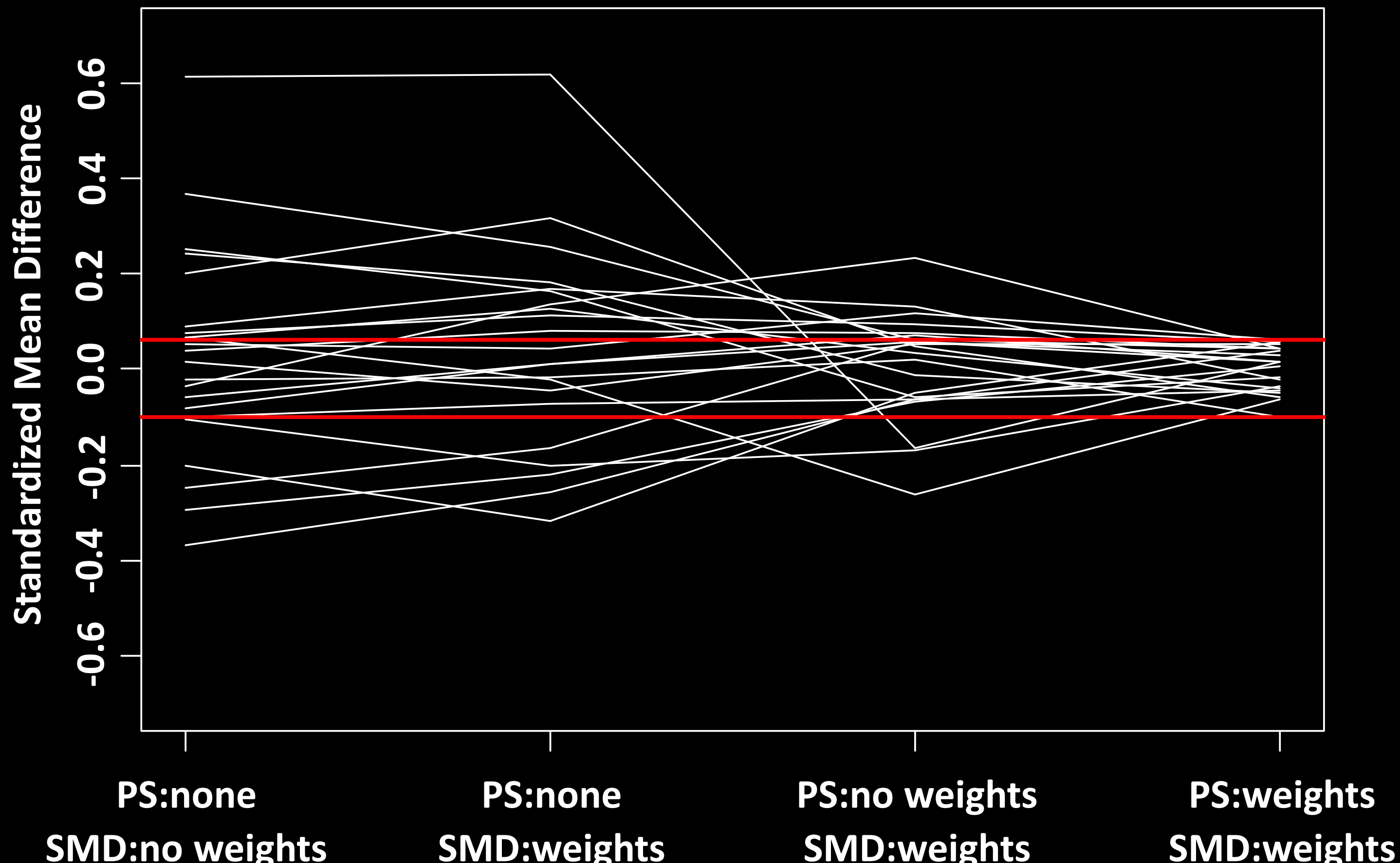
Using Sampling Weights Consistently Provides Lowest RMSE of ATT

		Standard	Random sample	Selection depends on (*x*, *t*)	Weight scales differ	Non-linear treatment
	Select	$s \sim x$	$s \perp x$	$s \sim (x, t)$	$s \sim x \| t$	$s \sim x$
	Treatment	$t \sim x$	$t \sim x$	$t \sim x$	$t \sim x$	$t \sim x^2$
Sampling weights	PS model					
No	$t \sim 1$	1.041	1.079	1.055	1.046	0.160
No	$t \sim x$	0.230	0.041	0.109	0.214	0.674
Covariate	$t \sim x + sw$	0.161	0.041	1.615	0.401	1.175
Weight	$t \sim x$	0.155	0.039	0.093	0.149	0.113

COPS Recruit Study Examined Barriers When Considering Policing Careers

- 2009 Insights from the Newest Members of America's Law Enforcement Community survey
- National pool of 1,600 respondents from 44 of the largest police and sheriff departments
- Asked recruits about
  - reasons for pursuing a career in law enforcement
  - disadvantages of such a career
  - influencers on career choices
  - perceived effectiveness of recruiting strategies
- Compare white and minority candidates on attractions and barriers
  - minority recruits more likely to be female, be married, have children, and never have attended college
  - respondents more likely to be married, have children, and never have attended college

Balance of 22 Covariates Improved by Using Sampling Weights Throughout



Effects Can Move Estimates in Either Direction

- Minority officers are significantly more concerned about benefits, particularly health insurance
  - odds ratio of 2.68 without sampling weights throughout
  - odds ratio of 2.56 with sampling weights throughout
- Police excessive force concerns minority officers to the point they consider not joining
  - odds ratio of 1.64 without sampling weights throughout
  - odds ratio of 1.93 with sampling weights throughout

Outline

- Background
- Derivation
- Risks that proper use of sampling weights avoid
- Simulation and analysis of police survey data
- Conclusions

Use Sampling Weights Throughout Propensity Score Analyses

- Estimate $P(t = 1|x)$ not $P(t = 1|x, s = 1)$, achievable by using sampling weights
- Weight treatment cases as $\frac{1}{P(s=1|t=1,\mathbf{x})}$
- Weight control cases as $\frac{1}{P(s=1|t=0,\mathbf{x})} \frac{P(t=1|\mathbf{x})}{1-P(t=1|\mathbf{x})}$
- Protects against
  - variables unavailable for propensity scores, but baked into sampling weights
  - allocating degrees of freedom to the wrong regions of **x**
  - sampling weights that depend on treatment assignment

Propensity Score Analysis with Survey Weighted Data

Greg Ridgeway
Department of Criminology
Department of Statistics

Stephanie A. Kovalchik
Victoria University
Game Intelligence Group, Tennis Australia

Beth Ann Griffin
RAND Corporation

Mohammed U. Kabeto
Department of Internal Medicine
University of Michigan

Appeared in the *Journal of Causal Inference*, 2015