Statistical Applications in Crime and Justice

Greg Ridgeway

Department of Criminology

Department of Statistics

Our Justice System is a Large Expenditure

Sector	Annual Expense
Health	$3000B
Finance/Insurance	$1000B
Defense	$500B
Agriculture/Food	$400B
Justice	$250B
Dentistry	$120B

Outline

- Natural experiments
- Propensity score/doubly robust methods
- Additional topics in criminology and statistics

Outline

- Natural experiments
  - Testing for racial bias in police stops
    - *Exercise*
  - Effect of sleep on aggression
  - Effect of transit on crime in LA
- Propensity score/doubly robust methods
- Additional topics in criminology and statistics

Outline

- Natural experiments
  - Testing for racial bias in police stops
    - *Exercise*
  - Effect of sleep on aggression
  - Effect of transit on crime in LA
- Propensity score/doubly robust methods
- Additional topics in criminology and statistics

Much of Criminology Depends on Natural Experiments

- Several well-known justice experiments
  - Milwaukee Domestic Violence RCT, late 1980s
  - Chicago Becoming a Man RCT, 2009-2015
- Natural experiments occur because of an external shock to the population being studied
  - Hurricane Katrina displaced thousands, re-incarceration rates were 15 percentage points lower for ex-prisoners forced to move to new communities
  - Discontinuities in COPS Office funding used to estimate the effect of police officers on crime

Outline

- Natural experiments
  - Testing for racial bias in police stops
    - *Exercise*
  - Effect of sleep on aggression
  - Effect of transit on crime in LA
- Propensity score/doubly robust methods
- Additional topics in criminology and statistics

Grogger & Ridgeway (2006). "Testing for racial profiling in traffic stops from behind a veil of darkness," *JASA* 101(475):878-887. ASA 2007 Outstanding Statistical Application Award

Racial Profiling Continues to Be a Concern

- I-95 "turnpike" studies in the mid-1990s raised public concern about racial profiling
- Public concern has led to state and local-level action
- Events periodically renew interest
  - Questionable police shootings
  - Arrest of Henry Louis Gates and the "beer summit" of July 2009

Unfortunately, the Quality of the Analysis Using Collected Data Is Weak

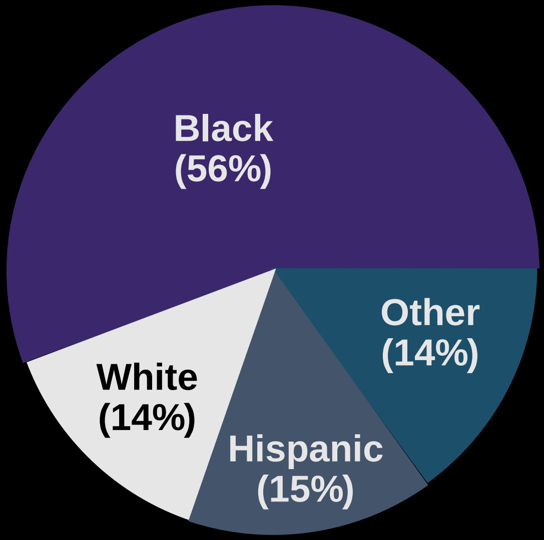
- A large number of studies claim racial profiling
  - Texas: Concluded that “75% of agencies stop more black and Latino drivers than white drivers”
- And some studies hastily conclude no profiling
  - Sacramento: The percentage of black drivers stopped matched the percentage of blacks among crime suspect descriptions

Why Is Testing for Racial Profiling So Hard?

Racial Distribution of People Stopped

Racial Distribution of People at Risk of Being Stopped

Difference Between



And

?

= Racial Profiling

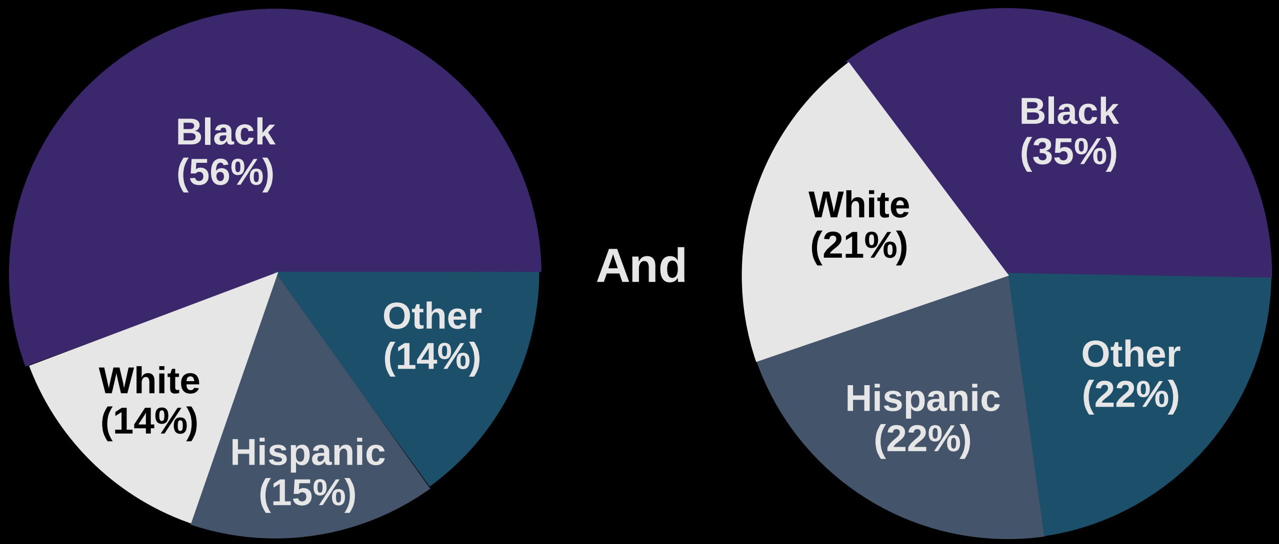
Source: Oakland Police Department, 2003

Why Is Testing for Racial Profiling So Hard?

Racial Distribution of People Stopped

Racial Distribution of Residents According to the Census

Difference Between



= 1.6

Source: Oakland Police Department, 2003

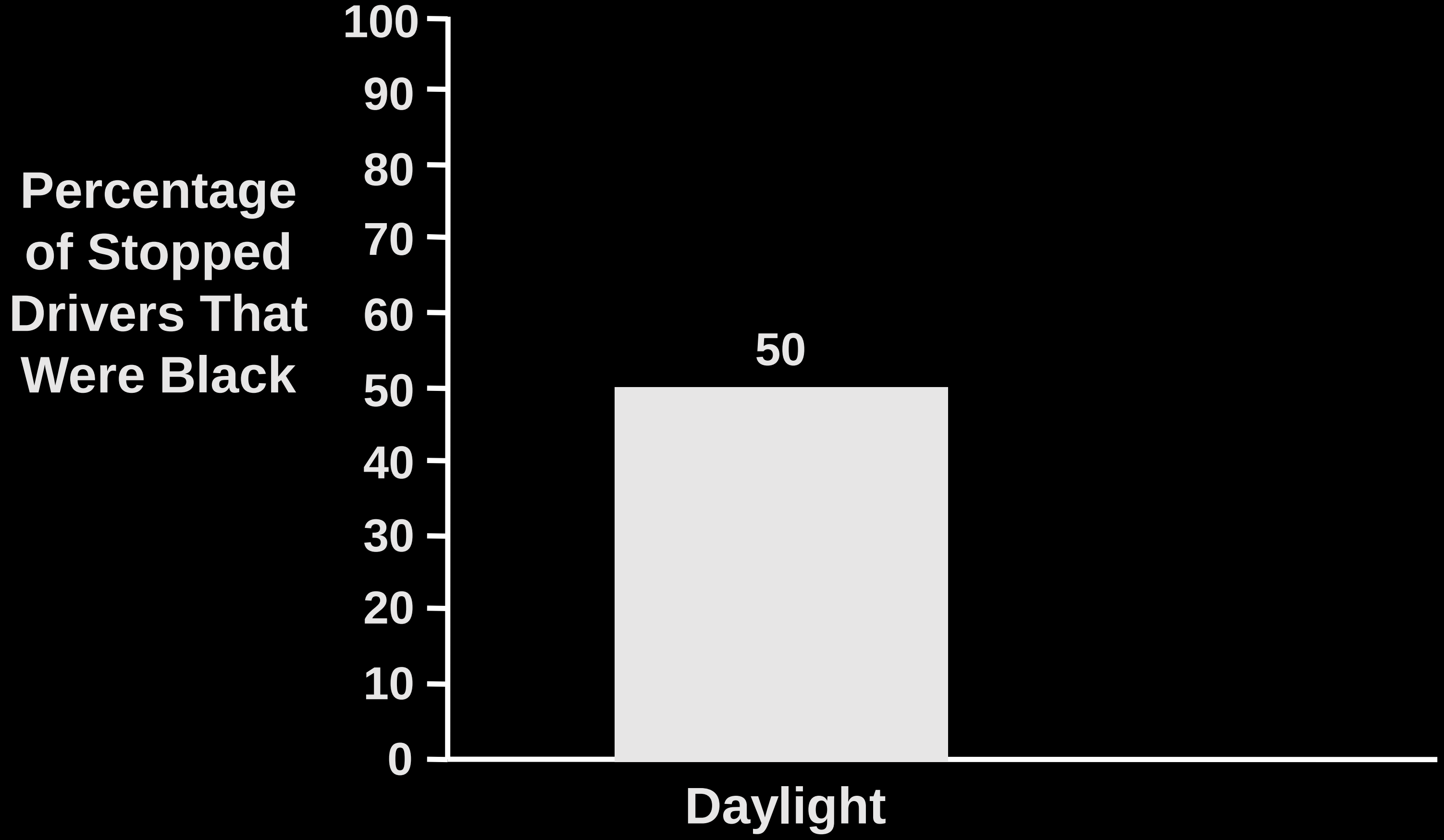
Source: U.S. Census, 2000

- The 1.6 disparity between the racial distributions may result from:
  - A race bias
  - Driving behavior: car ownership, time on the road, and care
  - Exposure to police by area of city, neighborhood characteristics, etc.

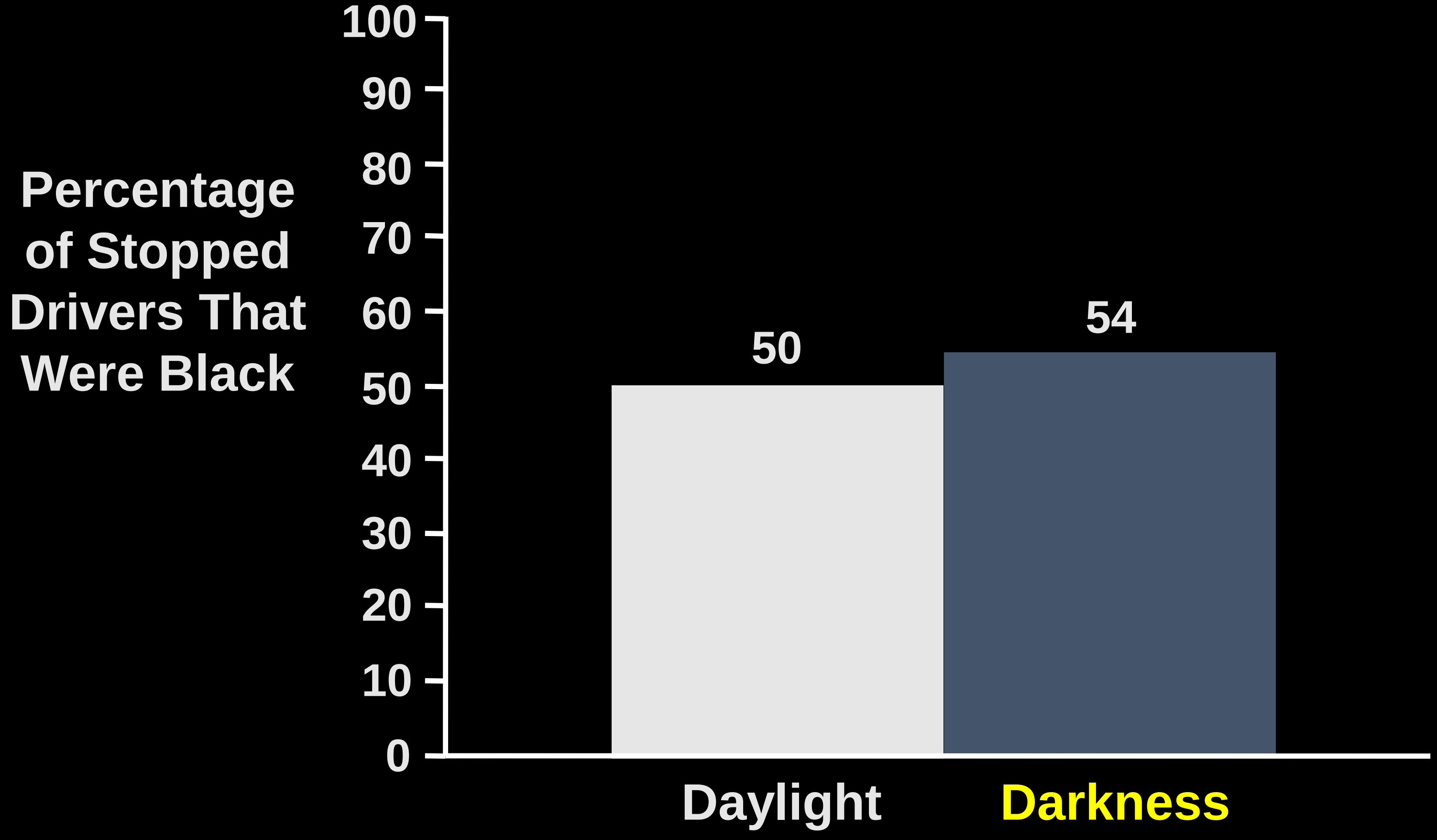
Does the Ability to See the Driver Influence Which Drivers Are Stopped?

- The ability to discriminate requires officers to identify the race in advance
- The ability to identify race in advance of the stop decreases as it becomes dark

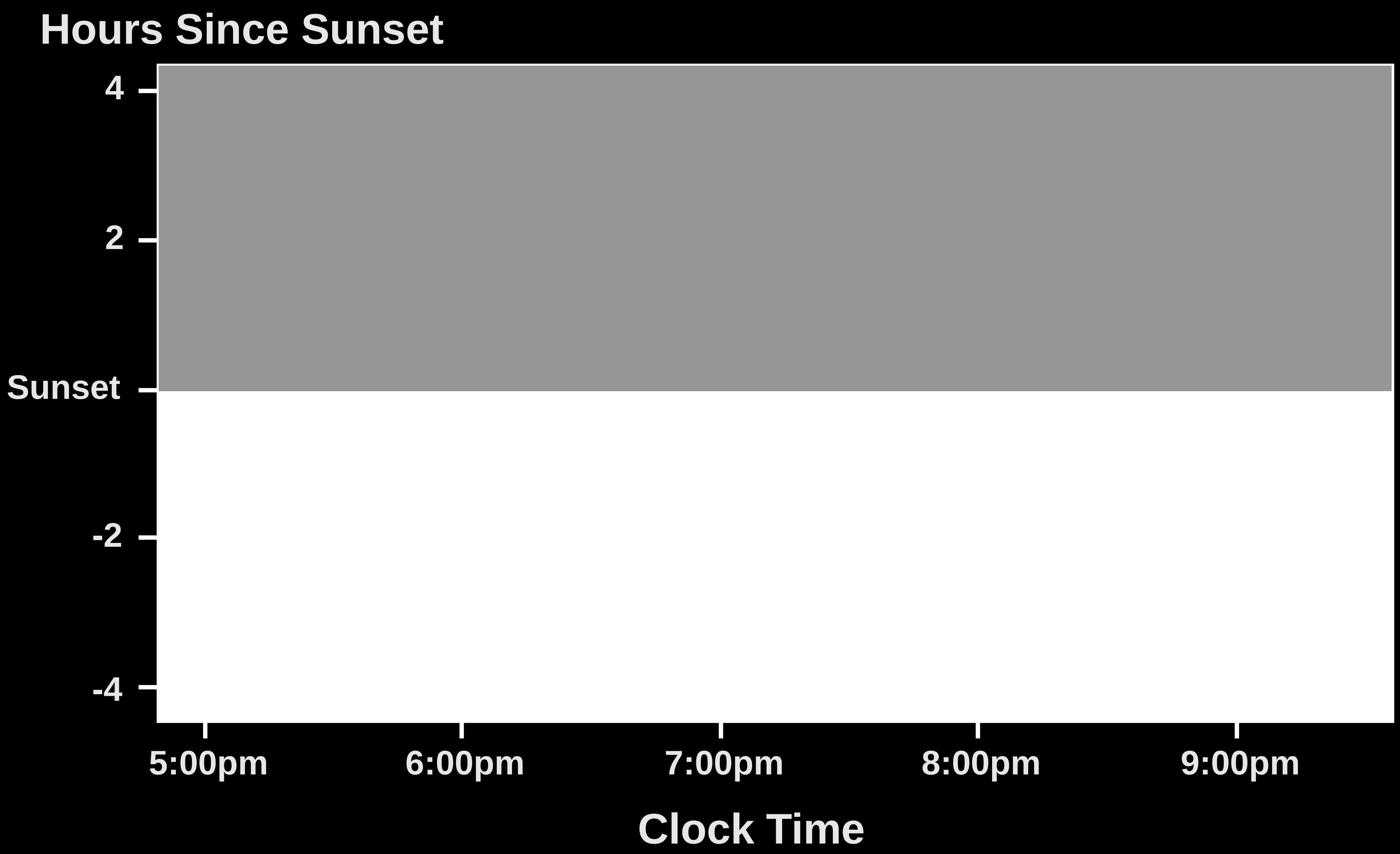
Simple "Veil of Darkness" Test Shows No Evidence of Racial Bias



Simple “Veil of Darkness” Test Shows No Evidence of Racial Bias

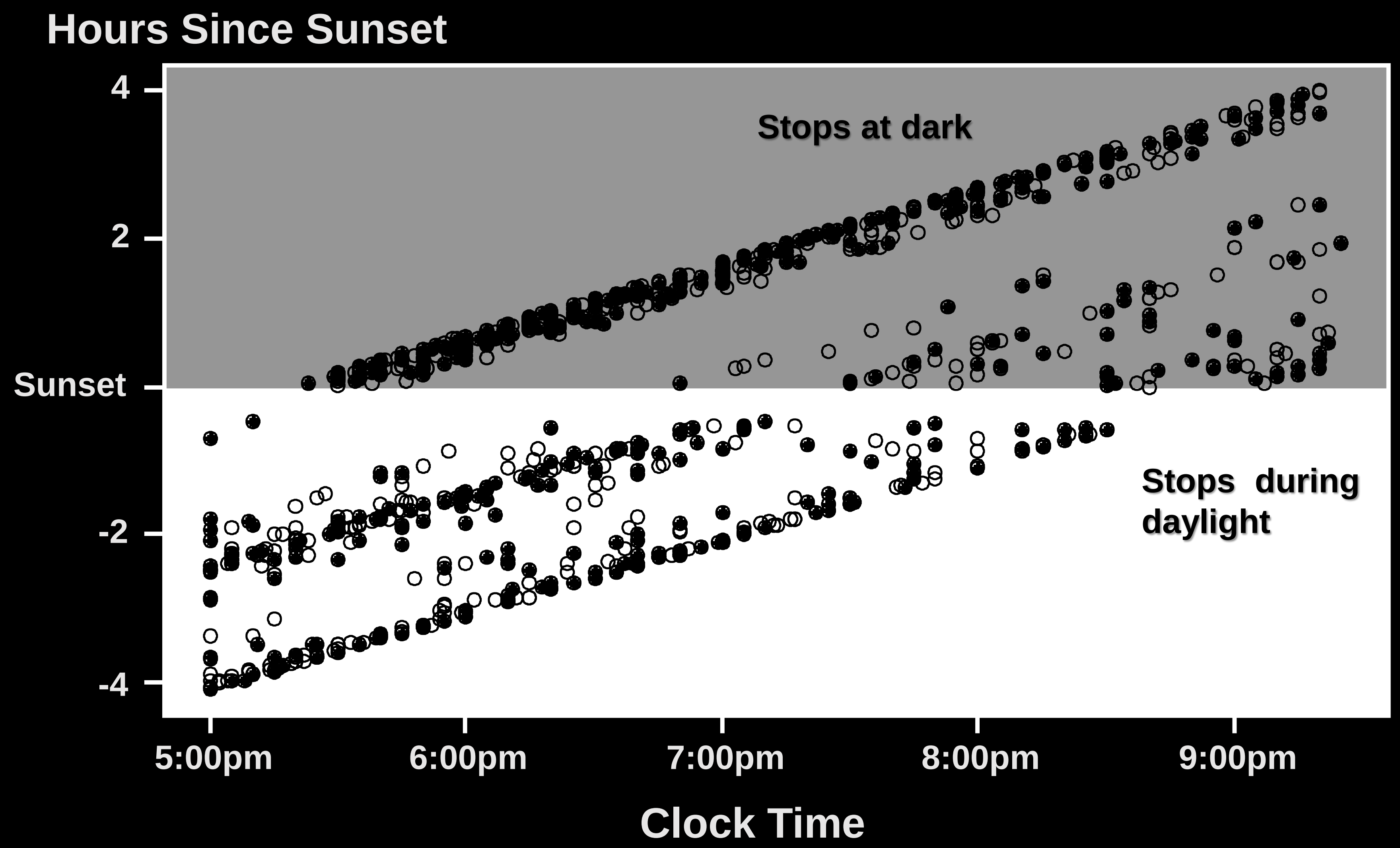


An Approach That Involved Adjusting for “Clock Time”



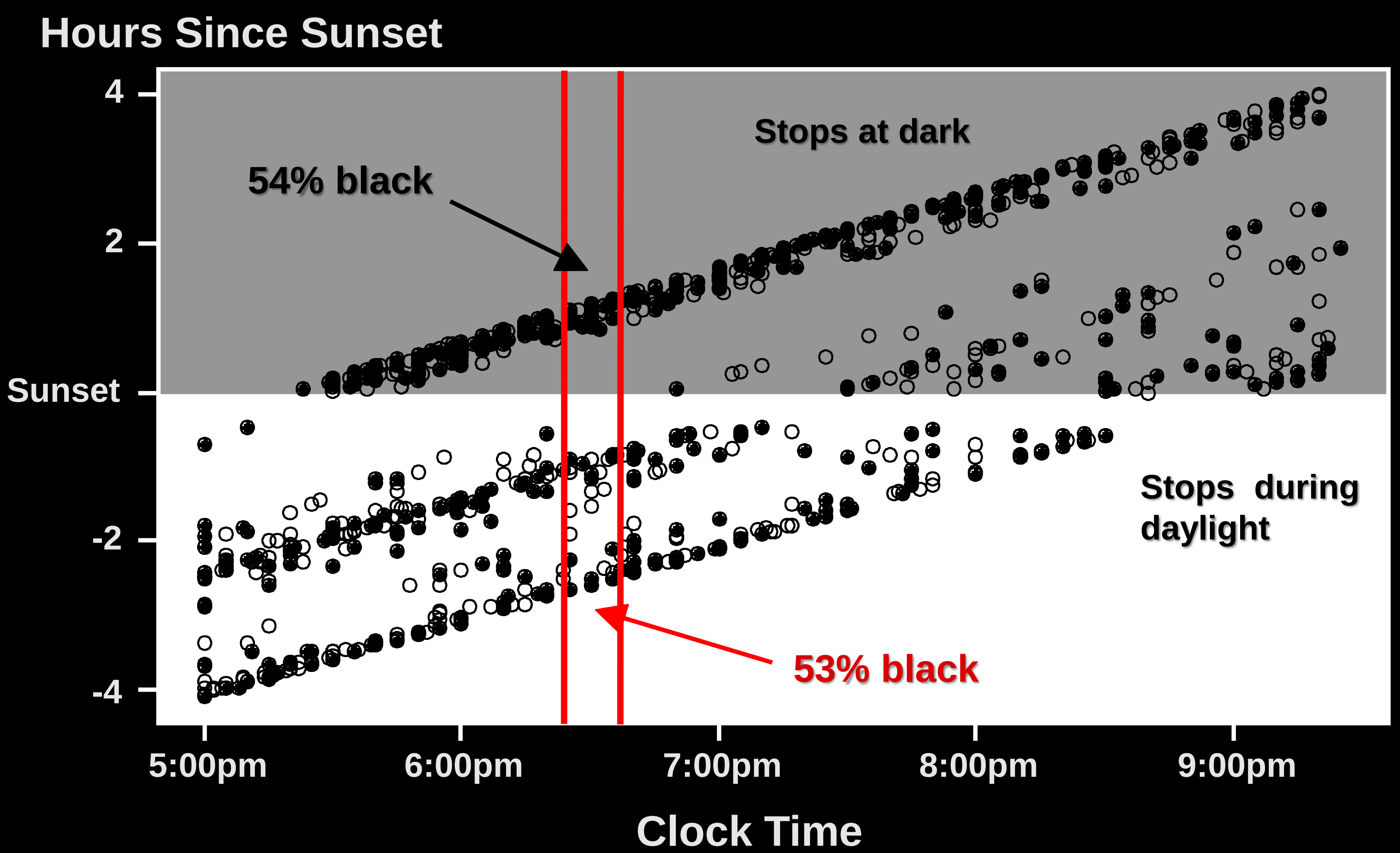
ASW Mar 2017

Compare Stops During Daylight with Stops in Darkness



ASW Mar 2017

There Is No Difference in the Rate that Black Drivers Are Stopped



ASW Mar 2017

Derivation of the VoD Estimator

$$\frac{P(S|B,V)}{P(S|\bar{B},V)}$$

- *S* – Stop
- *B* – Black driver
- *V* – Race is visible

Derivation of the VoD Estimator

$$\frac{P(S|B,V)}{P(S|\bar{B},V)} = K_{\text{ideal}} \frac{P(S|B,\bar{V})}{P(S|\bar{B},\bar{V})}$$

- S – Stop
- B – Black driver
- V – Race is visible

- $K_{\text{ideal}} > 1$ suggests officers are more likely to stop black drivers when their race is visible

Derivation of the VoD Estimator

$$\frac{P(S|B,t,d=0)}{P(S|\bar{B},t,d=0)} = K\frac{P(S|B,t,d=1)}{P(S|\bar{B},t,d=1)}$$

$$1 < K \leq K_{\text{ideal}}$$

- *S* – Stop
- *B* – Black driver
- *t* – Clock time
- *d* – Darkness

Assume

- $K_{\text{ideal}} > 1$
- $P(V|d=0) > P(V|d=1)$
- $\frac{P(B|d=1,t)}{P(\bar{B}|d=1,t)}\frac{P(\bar{B}|d=0,t)}{P(B|d=0,t)} = 1$

Decomposition of the VoD Estimator

$$K = \frac{P(B|R,S,t,d=0)}{1-P(B|R,S,t,d=0)} \frac{1-P(B|R,S,t,d=1)}{P(B|R,S,t,d=1)}$$

$$\frac{P(\bar{B}|t,d=0)}{P(B|t,d=0)} \frac{P(B|t,d=1)}{P(\bar{B}|t,d=1)}$$

$$\frac{P(R|\bar{B},S,t,d=0)}{P(R|\bar{B},S,t,d=1)} \frac{P(R|B,S,t,d=1)}{P(R|B,S,t,d=0)}$$


Dickies
BIKE LANE
PESCADO

VoD is Easily Implemented

- For each stop record race of driver, darkness indicator, and clock time
- Subset dataset to dates near the switch to/from Daylight Savings Time
- Logistic regression, predict race from darkness and clock time
- Report VoD estimate as $K = \exp(-\beta_1)$

Oakland 2003: $K = 0.88$
Cincinnati 2003-2008: $K = 0.96$

VoD Has Become Widely Used

- Connecticut
- San Diego
- Syracuse
- Urbana
- Minneapolis
- Raleigh-Durham

DURHAM NEWS MARCH 17, 2016 6:40 AM

RTI study finds racial disparities in Durham police traffic stops

HIGHLIGHTS

Among male drivers stopped, odds of driver being black were 20 percent higher during daylight

Study found racial disparities were greatest in specialized units; no disparity found in traffic unit

BY MARK SCHULTZ
AND THOMASI MCDONALD
mschultz@newsobserver.com
tmcdonald@newsobserver.com

DURHAM — A new analysis of Durham police statistics found that the proportion of black drivers stopped in recent years rose significantly during daylight hours when the race of the driver was presumably more apparent.

The analysis compared 151,701 traffic stops from January 2010 through October 2015 and found that the proportion of black drivers pulled over during daylight hours was 12 percent higher than the proportion during nighttime stops.

Among male drivers only, the odds that a driver was black were 20 percent higher when stopped during daylight than when stopped at night, according to the study by RTI International.

Outline

- Natural experiments
  - Testing for racial bias in police stops
    - *Exercise*
  - Effect of sleep on aggression
  - Effect of transit on crime in LA
- Propensity score/doubly robust methods
- Additional topics in criminology and statistics

Outline

- Natural experiments
  - Testing for racial bias in police stops
    - *Exercise*
  - Effect of sleep on aggression
  - Effect of transit on crime in LA
- Propensity score/doubly robust methods
- Additional topics in criminology and statistics

Umbach, Ridgeway, & Raine (under review). "Aggression and Lost Sleep: a Daylight Saving Time Natural Experiment"

Daylight Saving Time Generates Many Natural Experiments

- Lahti et al. (2006) found the spring DST change decreases sleep by 60.14 minutes
- DST natural experiments have found
  - Rise in car accidents (Harrison 2013)
  - Stock market losses (Kamstra, et al 2000 Pinegar 2002)
  - Work-place injuries (Barnes and Wagner 2009)
  - Cyberloafing (Wagner et al 2012)
  - Reduced test scores (Gaski and Sagarin 2011)
  - Crime due to less lighting (Doleac and Sanders 2015)
  - Suicide rates (Berk et al 2008)

NIBRS and City Assault Data Indicate Spring Decrease, Fall Rise

	Spring		Fall	
Year	Monday immediately following DST	Second Monday after start of DST	Monday immediately following DST	Second Monday after start of DST
2001	1797	1827	1673	1593

NIBRS and City Assault Data Indicate Spring Decrease, Fall Rise

	Spring		Fall	
Year	Monday immediately following DST	Second Monday after start of DST	Monday immediately following DST	Second Monday after start of DST
2001	1797	1827	1673	1593
2002	1652	2015	1671	1625

NIBRS and City Assault Data Indicate Spring Decrease, Fall Rise

	Spring		Fall	
Year	Monday immediately following DST	Second Monday after start of DST	Monday immediately following DST	Second Monday after start of DST
2001	1797	1827	1673	1593
2002	1652	2015	1671	1625
2003	1695	2032	1732	1951
2004	3597	3512	4019	3458
2005	2396	2355	Not Included Due to Halloween	
2006	2624	2766	2645	2430
2007	2681	2556	2397	2432
2008	Not Included Due to St. Patrick's Day		2639	2215
2009	2862	2832	2590	2741
2010	2732	2714	2564	2575
2011	2633	2932	2620	2841
2012	2873	2956	2541	2491
2013	2464	2635	2515	2331
2014	Not Included Due to St. Patrick's Day		2586	2607
Total	61,138		63,482	

Spring DST Reduces Assaults 3%

$$y_{itd} \sim \text{Poisson}(\lambda_{itd})$$
$$\lambda_{itd} = \beta d + \alpha_i + \delta_t$$

- $100(e^{\beta} - 1)$ gives percent increase in crime attributable to switch to DST
- -3.0% (-4.3%, -1.6%)

Falsification Tests Indicate That DST Effect is Genuine

Falsification tests check for effects unlikely to be causally related to the intervention

- Energy Policy Act of 2005 moved DST start date about four weeks earlier starting in 2007
  - Swapped the DST coding before and after 2007
  - 0.6% (-0.8%, 1.9%)
- Compared 1 and 2 weeks after DST
  - 0.3% (-1.1%, 1.8%)
- Compared Wednesdays
  - -0.3% (-1.6%, 0.9%)
- Compared Thursdays
  - -1.2% (-2.4%, 0.1%)

Fall Change to Standard Time Has an Unclear Impact on Assault

- Monday after to Monday 1 week later
  - 2.6% (1.3%, 4.0%)
- Energy Policy Act of 2005 recoding
  - 0.8% (-0.6%, 2.2%)
- Compared 1 and 2 weeks after DST
  - 2.6% (1.3%, 3.9%)
- Compared Wednesdays
  - 4.3% (2.9%, 5.6%)
- Compared Thursdays
  - 2.1% (0.8%, 3.5%)

Outline

- Natural experiments
  - Testing for racial bias in police stops
    - *Exercise*
  - Effect of sleep on aggression
  - Effect of transit on crime in LA
- Propensity score/doubly robust methods
- Additional topics in criminology and statistics

Ridgeway & MacDonald (2017, to appear). "Effect of Rail Transit on Crime: A Study of Los Angeles from 1988-2014," *Journal of Quantitative Criminology*

Belief that Transit Brings Crime

- Prompted by the extension of Metro Rail to Santa Monica, California
- Plano (1993) compared crime in the year before and year after three stations opened in Baltimore, compared to the rest of Baltimore County
  - Found no effect
- Poister (1996) examined two stations in Atlanta 2½ years before and 1½ years after opening
  - Found no effect
- Block and Block (2000) found higher robbery rates around 1-2 blocks away from transit stops in Chicago and New York
  - also more likely to be located near bars and other businesses that may be sources of crime
- Ihlanfeldt (2003) studied transit expansion in Atlanta from 1991-1994, crime increased near downtown and decreased in the suburbs
- Liggett et al. (2003) studied 14 new Metro Rail "Green Line" stations connecting poor neighborhoods to more affluent neighborhoods
  - compared the crime rates for the five years before and after opening relative to the local city or larger jurisdiction in which each station was located
  - Found crimes increase in six out of the fourteen station areas relative to the adjacent areas in which each station was situated

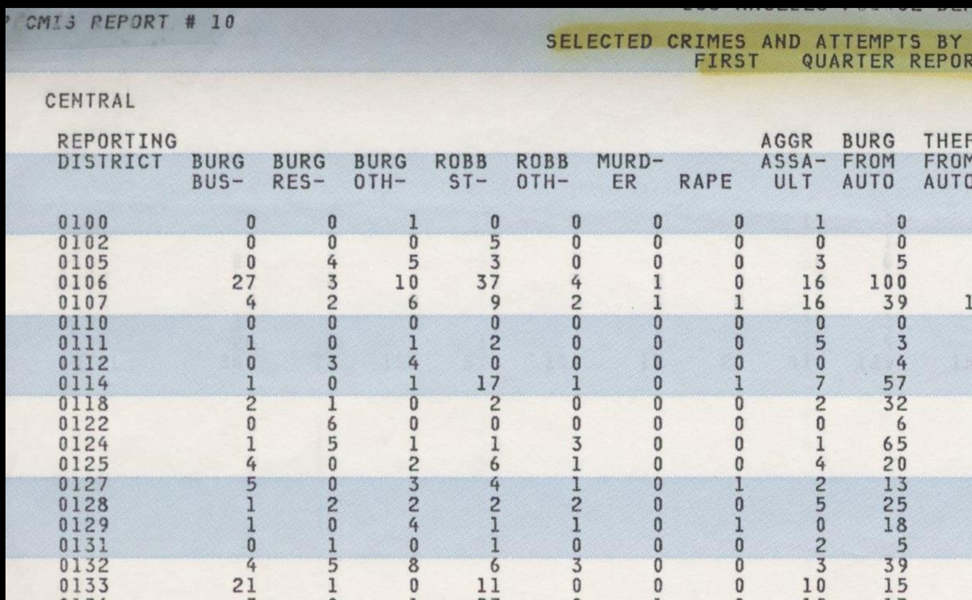
Though Best Known for Freeways, Los Angeles Has Built an Extensive Rail System

- 1961 the last of the Pacific Electric rail lines ended service
- 1961-1990 Los Angeles was the largest city in the U.S. without a rail transit system
- 1990 Los Angeles opened the Blue Line
- 2014 Los Angeles had six lines covering 87 miles of service, carrying more than 300,000 daily riders

Los Angeles Presents a Special Opportunity to Assess the Effect of Transit on Crime

1. Data on crime trends cover the entire expansion of Metro Rail in the second largest U.S. city
2. The time series is nearly three decade long
3. Compare crime near stations before and after opening and with areas eventually having stations
4. Transit labor union went on a 32-strike in 2000 and a 35-day strike in 2003

Data From 1988-2004 Collected from 2,300 Pages at LA Library

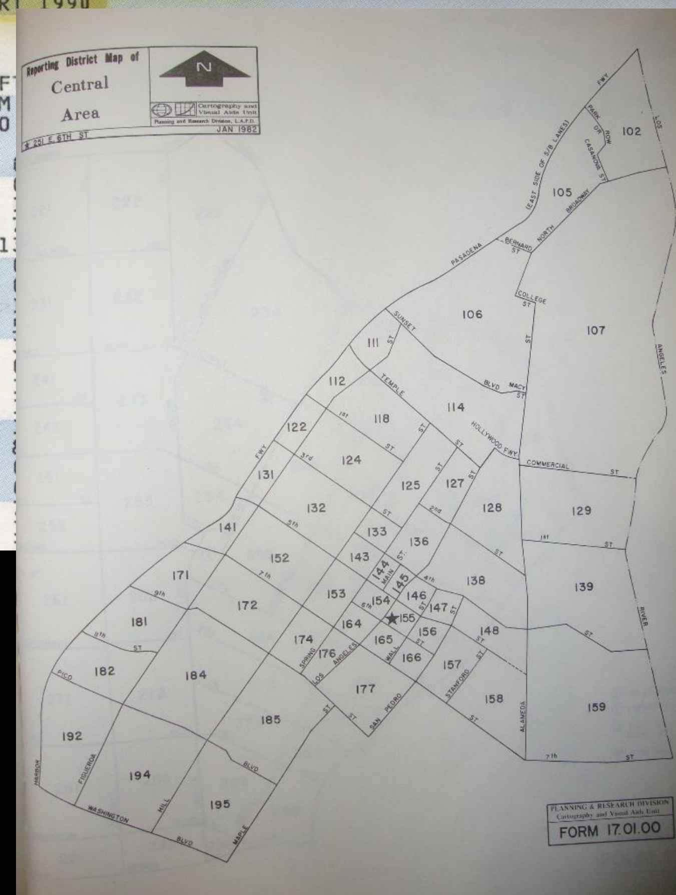


CMIS REPORT # 10

SELECTED CRIMES AND ATTEMPTS BY REPORTING DISTRICT
FIRST QUARTER REPORT 1990

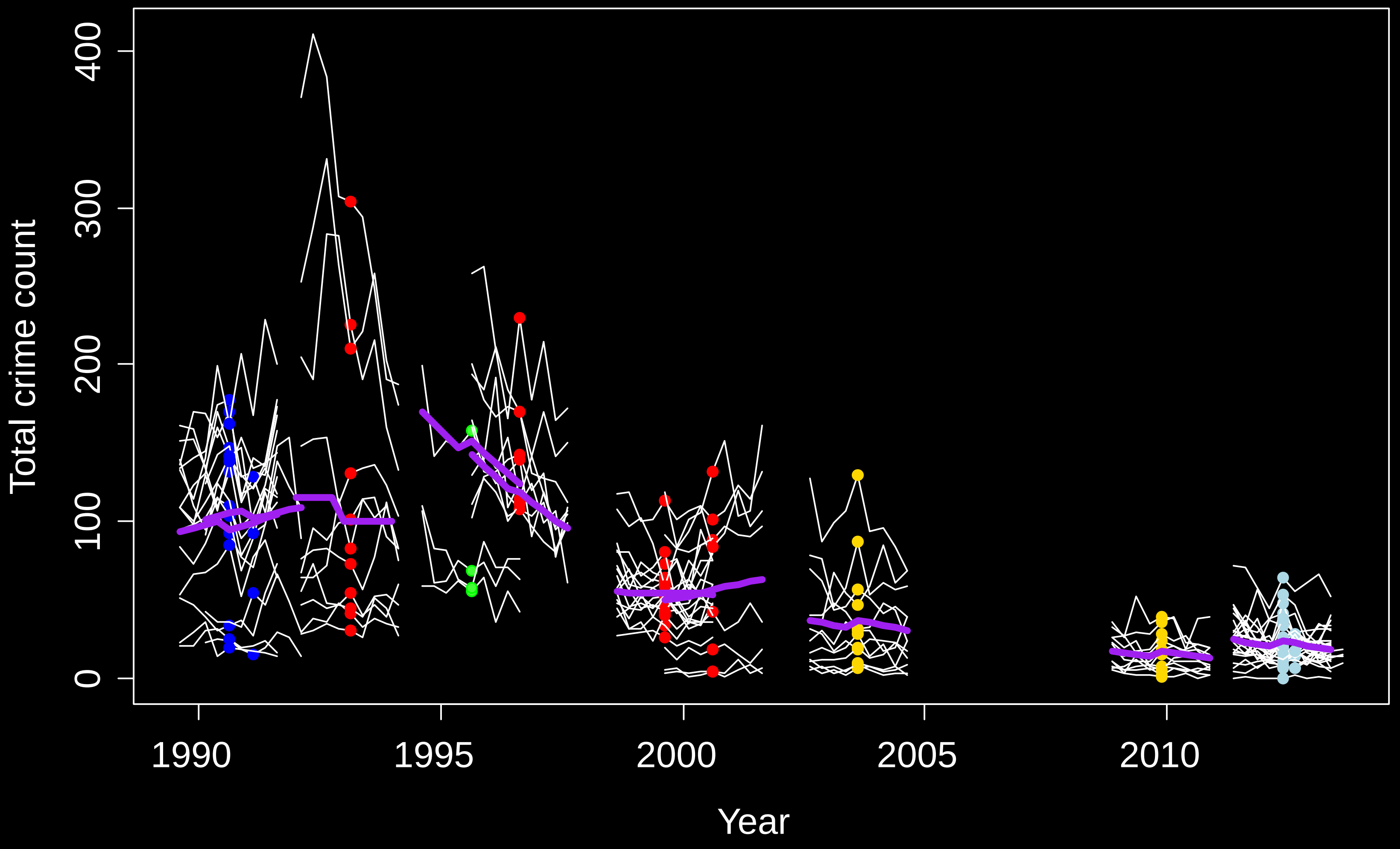
CENTRAL

REPORTING DISTRICT	BURG BUS-	BURG RES-	BURG OTH-	ROBB ST-	ROBB OTH-	MURD-ER	RAPE	AGGR ASSA-ULT	BURG FROM AUTO	
0100	0	0	1	0	0	0	0	1	0	3
0102	0	0	0	5	0	0	0	0	0	7
0105	0	4	5	3	0	0	0	3	5	26
0106	27	3	10	37	4	1	0	16	100	17
0107	4	2	6	9	2	1	1	16	39	49
0110	0	0	0	0	0	0	0	0	0	1
0111	1	0	1	2	0	0	0	5	3	22
0112	0	3	4	0	0	0	0	0	4	15
0114	1	0	1	17	1	0	1	7	57	20
0118	2	1	0	2	0	0	0	2	32	56
0122	0	6	0	0	0	0	0	0	6	14
0124	1	5	1	1	3	0	0	1	65	96
0125	4	0	2	6	1	0	0	4	20	59
0127	5	0	3	4	1	0	1	2	13	60
0128	1	2	2	2	2	0	0	5	25	69
0129	1	0	4	1	1	0	1	0	18	36
0131	0	1	0	1	0	0	0	2	5	15
0132	4	5	8	6	3	0	0	3	39	20
0133	21	1	0	11	0	0	0	10	15	74



- Data from 2005-2014 came from LAPD incident level crime data
- All data available at github.com/gregridgeway/LAPDcrimedata

Identification Strategy Relies on the Staged Rollout of Metro Rail over 30 Years



ASW Mar 2017

Isolate Effect of Transit Using Four Approaches

- *Stepped wedge design* – compare RDs with and without stations over time
- *Effect modification* – measure the effect for “high crime” and “low crime” RDs
- *Short term, station RDs only* – analyze only RDs with stations in the year before and after a station
- *Two labor strikes* – use the strikes to assess crime changes before, during, after the system shutdown

Isolate Effect of Transit Using Four Approaches

- *Stepped wedge design* – compare RDs with and without stations over time
- *Effect modification* – measure the effect for "high crime" and "low crime" RDs
- *Short term, station RDs only* – analyze only RDs with stations in the year before and after a station
- *Two labor strikes* – use the strikes to assess crime changes before, during, after the system shutdown

Stepped Wedge Design Detects Shifts in Crime Rates When Stations Open

$$\log(\lambda_{it}) = \beta_0 + \beta_1 \text{station}_{it} + \alpha_i + \gamma'_{\text{division}(i)} ns_{15}(t)$$

- RDs have a station if the station is within 200m of the RD boundary
- 281 RDs within 1km of an eventual station
- 116 RDs will eventually have a station
- α_i is the RD fixed effect
- Allow for a separate crime trend in each division across the 108 quarters
- Computed permutation p-values by randomly exchanging station openings between RDs

Slight, Non-Significant Decline in Crime After Station Opening

Crime type	Average crime count per RD per year	% crime increase	95% CI	Permutation p-value
Total	216.9	-2.6	(-6.2, 1.2)	0.21

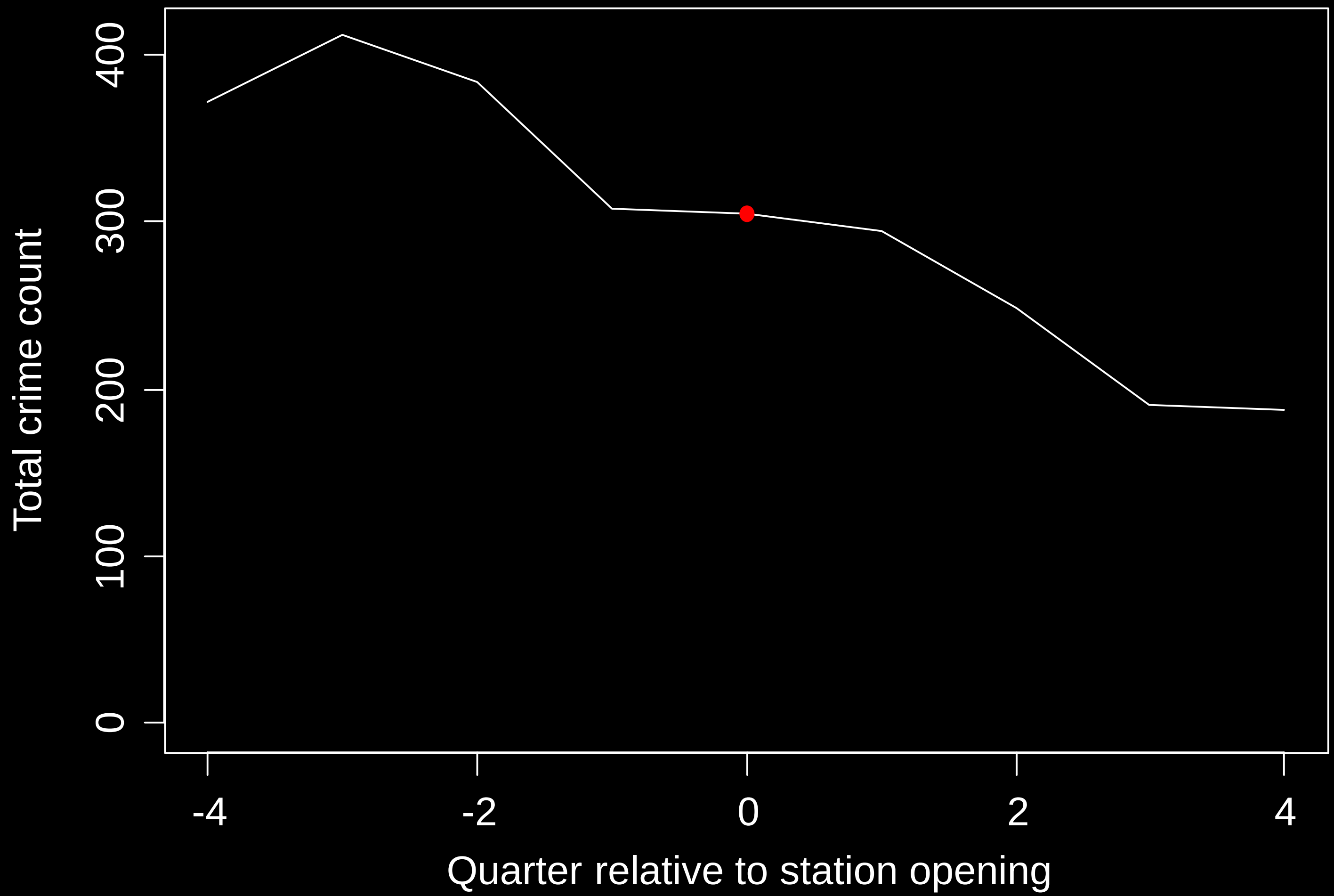
Most Crime Types Decline After Station Opening, None Significant

Crime type	Average crime count per RD per year	% crime increase	95% CI	Permutation p-value
Total	216.9	-2.6	(-6.2, 1.2)	0.21
Assaults	39.8	-3.9	(-9.4, 1.9)	0.17
Burglary/theft from vehicle	58.3	-3.6	(-9.9, 3.1)	0.34
Burglary	34.5	-2.2	(-8.1, 4.2)	0.51
Auto theft	46.1	-3.8	(-9.3, 2.0)	0.17
Grand theft person	4.5	-6.9	(-19.0, 7.1)	0.35
Homicide	0.9	4.6	(-8.6, 19.7)	0.51
Robbery	32.8	-0.9	(-7.9, 6.6)	0.77

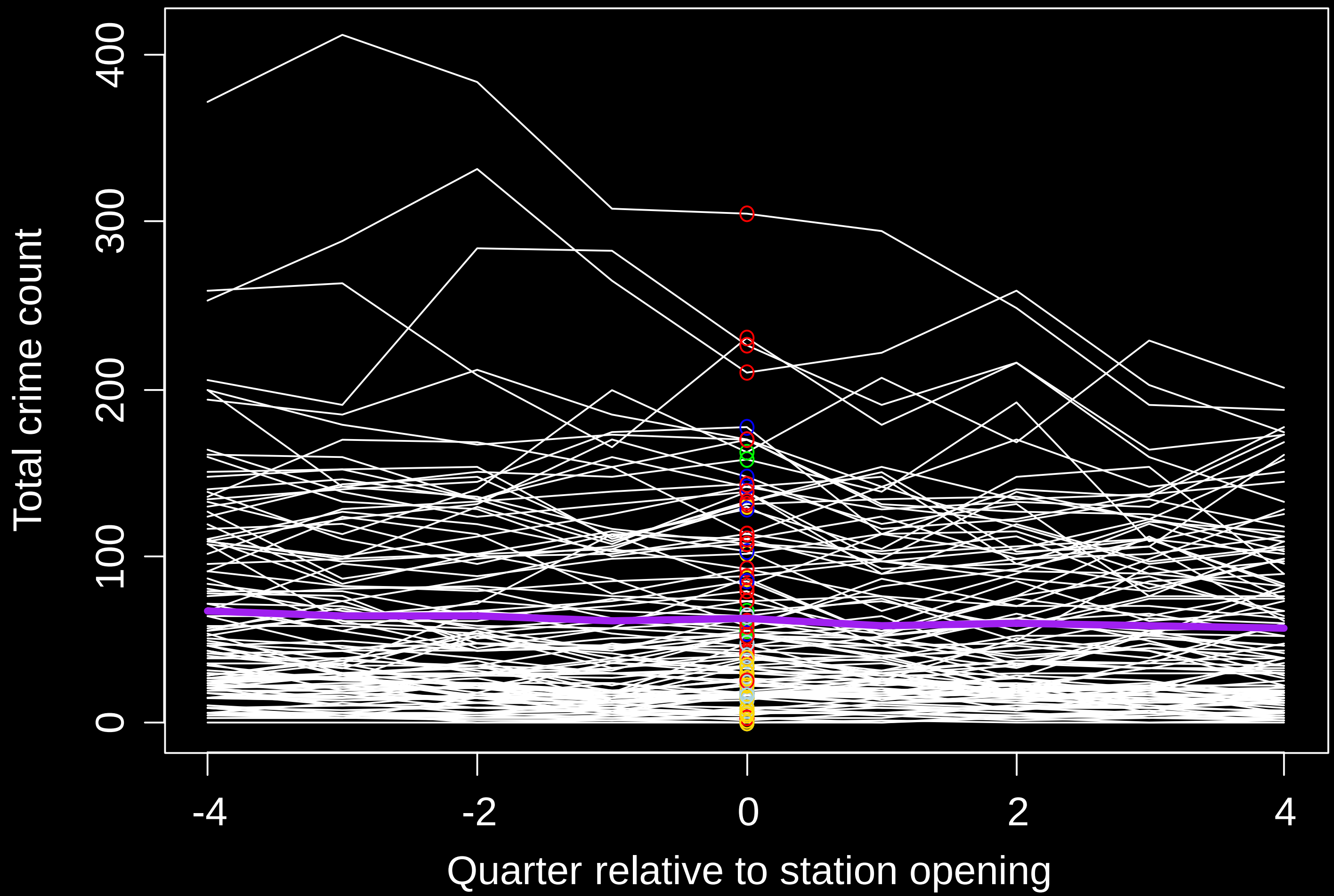
Isolate Effect of Transit Using Four Approaches

- *Stepped wedge design* – compare RDs with and without stations over time
- *Effect modification* – measure the effect for “high crime” and “low crime” RDs
- *Short term, station RDs only* – analyze only RDs with stations in the year before and after a station
- *Two labor strikes* – use the strikes to assess crime changes before, during, after the system shutdown

Examining Only Station RDs Avoids Confounding of Opening and Crime

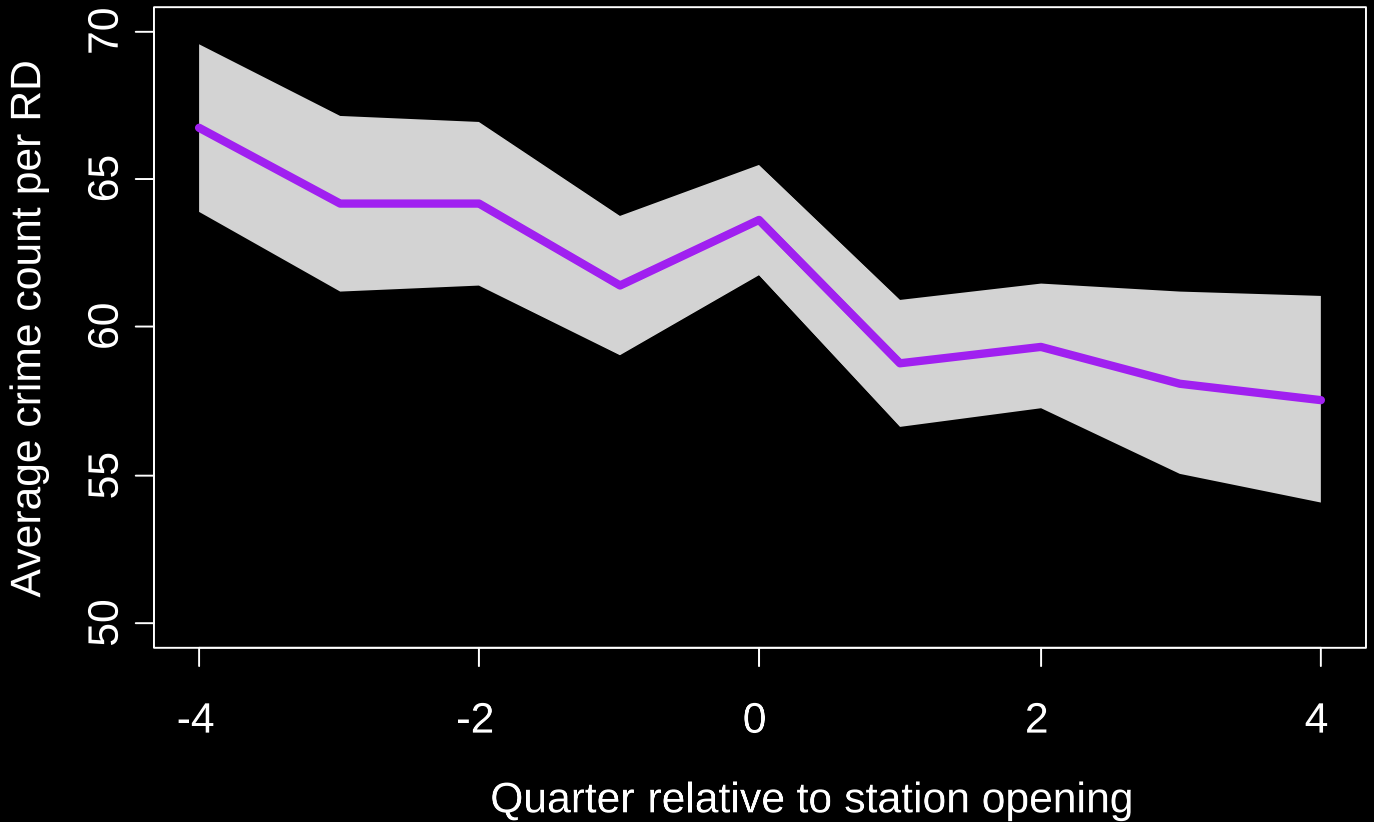


Examining Only Station RDs Avoids Confounding of Opening and Crime



ASW Mar 2017

3% Increase When Station Opens, But Could Be Random



Station Openings Have a Minimal Effect on Crime

Crime type	% increase in crime when station opens	95% CI	p-value			
Total	2.7	(-3.1, 8.9)	0.43			

P-values calculated by randomly selecting a different nine quarter sequence from the same RDs

Effect of Station Opening Does Not Vary by Distance to Station

Crime type	% increase in crime when station opens	95% CI	p-value	% crime increase per km away from station	95% CI	p-value
Total	2.7	(-3.1, 8.9)	0.43	0.1	(-0.1, 0.4)	0.50

P-values calculated by randomly selecting a different nine quarter sequence from the same RDs

No Strong Relationship Between Station Opening and Distance to Station

Crime type	% increase in crime when station opens	95% CI	p-value	% crime increase per km away from station	95% CI	p-value
Total	2.7	(-3.1, 8.9)	0.43	0.1	(-0.1, 0.4)	0.50
Assaults	-0.6	(-9.9, 9.6)	0.91	0.0	(-0.5, 0.6)	0.97
Burglary/theft from vehicle	5.6	(-3.0, 14.8)	0.40	0.0	(-0.4, 0.4)	0.95
Burglary	1.5	(-13.1, 18.5)	0.85	0.5	(-0.2, 1.2)	0.26
Auto theft	6.6	(-5.8, 20.7)	0.23	0.1	(-0.3, 0.5)	0.83
Grand theft person	-8.7	(-28.6, 16.7)	0.51	1.2	(0.2, 2.3)	0.18
Homicide	-27.7	(-59.9, 30.4)	0.31	0.0	(-3.5, 3.5)	0.98
Robbery	0.6	(-9.0, 11.3)	0.92	0.1	(-0.6, 0.7)	0.88

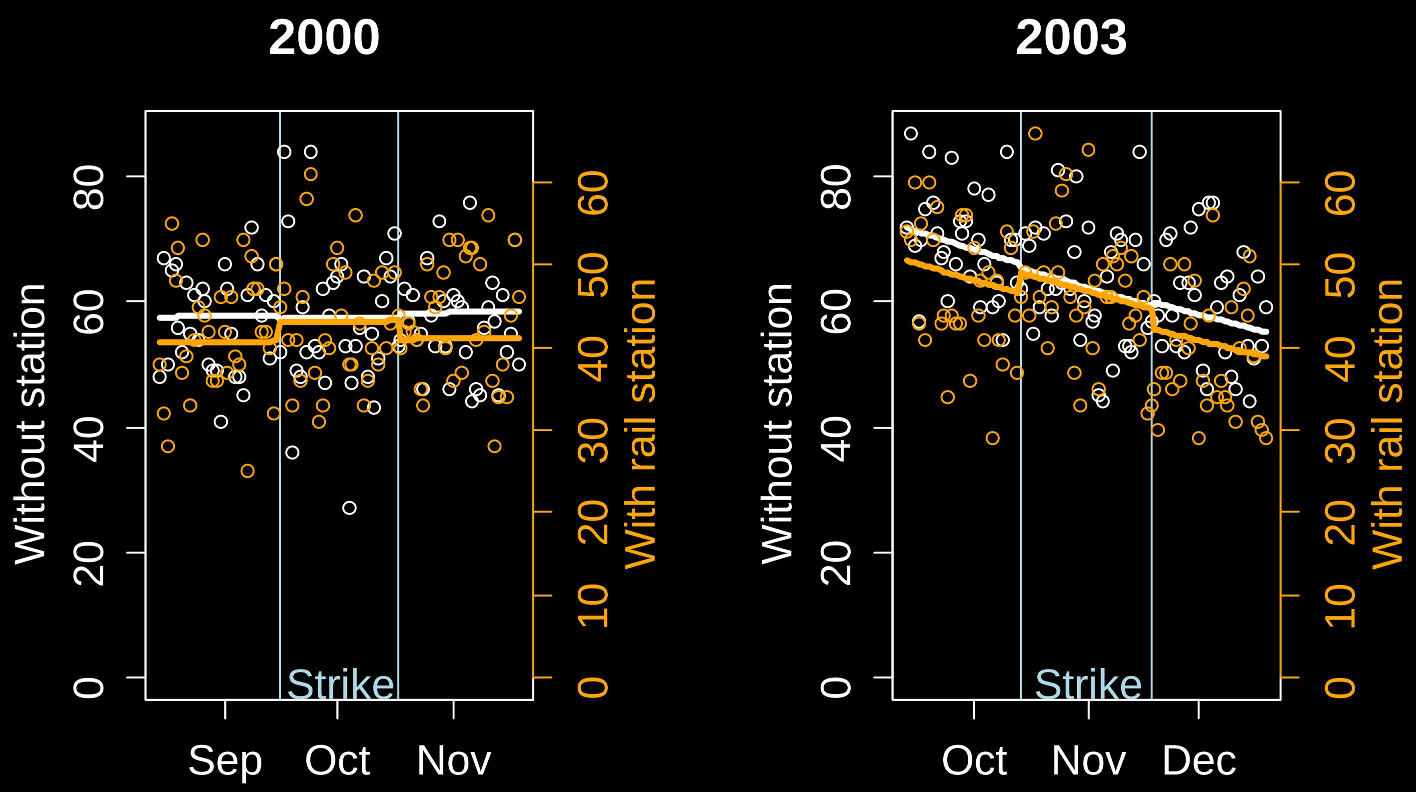
Isolate Effect of Transit Using Four Approaches

- *Stepped wedge design* – compare RDs with and without stations over time
- *Effect modification* – measure the effect for "high crime" and "low crime" RDs
- *Short term, station RDs only* – analyze only RDs with stations in the year before and after a station
- *Two labor strikes* – use the strikes to assess crime changes before, during, after the system shutdown

Strikes in 2000 and 2003 Shutdown Los Angeles Transit System

- 32-day strike from September 16-October 17, 2000
- 35-day strike ran from October 14-November 18, 2003
- Lo and Hall (2006) and Anderson(2014) showed strikes substantially disrupted transit

Crime Appears to Increase Near Stations During the Transit Strikes



Compare Strike and Non-strike Periods and Transit and Non-transit RDs

$$
\begin{aligned}
\log(\lambda_{it}) = \beta_0 + \beta_1 \text{transit}_i + \beta_2 \text{strike}_t + \\
\beta_3 \text{transit}_i \text{strike}_t + \\
\beta_4 I(\text{year}(t) = 2003) + \\
\beta_5 I(\text{year}(t) = 2000) t + \\
\beta_6 I(\text{year}(t) = 2003) t
\end{aligned}
$$

$$
\exp(\beta_3) = \frac{\dfrac{\lambda_{\text{transit},\text{strike}}}{\lambda_{\text{transit},\overline{\text{strike}}}}}{\dfrac{\lambda_{\overline{\text{transit}},\text{strike}}}{\lambda_{\overline{\text{transit}},\overline{\text{strike}}}}}
$$

Total Crime and Theft From Vehicles Increase During Strike

Crime type	Relative increase in crime at stations during strike	95% CI	Permutation p-value
Total	1.07	(0.99, 1.14)	0.068
Assaults	1.02	(0.87, 1.21)	0.79
Burglary/theft from vehicle	1.10	(0.98, 1.24)	0.15
Burglary	1.13	(0.95, 1.34)	0.18
Auto theft	1.06	(0.92, 1.21)	0.46
Grand theft person	1.06	(0.66, 1.70)	0.83
Homicide	2.22	(0.76, 6.51)	0.17
Robbery	1.00	(0.84, 1.19)	0.99

Public Transit Has Numerous Benefits, Neither Promotes Nor Hinders Crime

- Expansion of public transit has been justified as a basis for
  - reducing traffic congestion
  - improving economic development, and
  - reducing spatial mismatch of employment and low income households
- Neighborhoods often resist public transit expansion for fears that it will increase crime in neighborhoods
  - easier for criminals to travel to wealthier neighborhoods
  - increase the number of transient people to areas, generating signs of disorder
  - more potential victims traveling in relatively unguarded environments
- We find no impact of Metro Rail expansion on crime, positive or negative
  - Suggests crime should not be a factor for or against transit expansion

When We Return…

- Natural experiments
- Propensity score/doubly robust methods
  - Race bias in post-stop outcomes
    - *Exercise*
  - Performance benchmarking officers and communities
- Additional topics in criminology and statistics

Outline

- Natural experiments
- Propensity score/doubly robust methods
    - Race bias in post-stop outcomes
        - *Exercise*
    - Performance benchmarking officers and communities
- Additional topics in criminology and statistics

Ridgeway (2006). "Assessing the effect of race bias in post-traffic stop outcomes using propensity scores," *Journal of Quantitative Criminology* 22(1):1-29

McCaffrey, Ridgeway, & Morral (2004). "Propensity score estimation with boosted regression for evaluating adolescent substance abuse treatment," *Psychological Methods* 9(4):403-425

Post Stop Outcomes Provide an Opportunity to Assess Racial Bias

- Auditing police-citizen interactions
    - Video taped analysis
- Hit Rates
    - Comparing yields from contraband searches
- Matching on characteristics of stopped citizens
    - Comparing race groups who are similarly situated
    - Use the same methodology for matching officers' stops

Cincinnati Reported Large Disparities in Stop Duration

Stop feature	% Black drivers N=26,941		% Nonblack drivers (unadjusted) N=25,149
Stop < 10 minutes	55		65

Black and Nonblack Drivers Differ in Numerous Ways

Stop feature	% Black drivers N=26,941		% Nonblack drivers (unadjusted) N=25,149
Stop < 10 minutes	55		65
Invalid license	22		7
Male	65		66
Neighborhood			
Over-the-Rhine	9		5
Avondale	5		1
I-75	4		11
Residence			
Cincinnati	93		61
Date\Time			
12am-4am	16		8
Monday	15		14
August	9		11
Age			
18-25	33		29
Reason			
Equipment violation	27		16
Moving violation	51		76

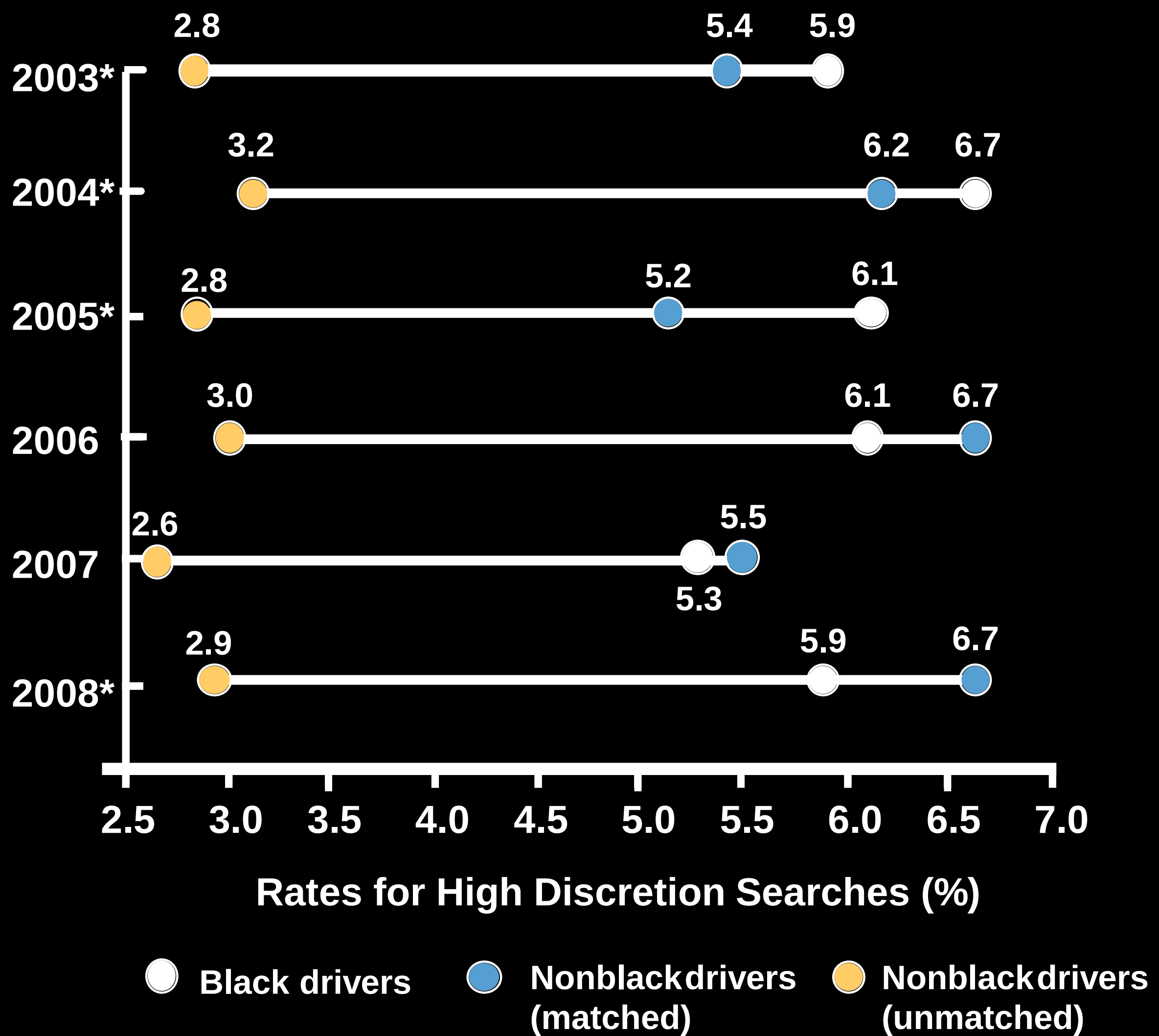
There Are Similarly Situated Nonblack Drivers

Stop feature	% Black drivers N=26,941	% Nonblack drivers (weighted) ESS=4,952	% Nonblack drivers (unadjusted) N=25,149
Stop < 10 minutes	55		65
Invalid license	22	20	7
Male	65	65	66
Neighborhood			
Over-the-Rhine	9	10	5
Avondale	5	5	1
I-75	4	5	11
Residence			
Cincinnati	93	92	61
Date\Time			
12am-4am	16	16	8
Monday	15	15	14
August	9	9	11
Age			
18-25	33	32	29
Reason			
Equipment violation	27	28	16
Moving violation	51	52	76

No Difference in Stop Duration

Stop feature	% Black drivers N=26,941	% Nonblack drivers (weighted) ESS=4,952	% Nonblack drivers (unadjusted) N=25,149
Stop < 10 minutes	55	57	65
Invalid license	22	20	7
Male	65	65	66
Neighborhood			
Over-the-Rhine	9	10	5
Avondale	5	5	1
I-75	4	5	11
Residence			
Cincinnati	93	92	61
Date\Time			
12am-4am	16	16	8
Monday	15	15	14
August	9	9	11
Age			
18-25	33	32	29
Reason			
Equipment violation	27	28	16
Moving violation	51	52	76

Most of the Search Rates Disparity Is Also Due to Non-Racial Factors



Simulated Example Shows Problems with Standard Regression Approach

	Treatment	Control
y	0.89	5.02
x_1	27%	90%
x_2	30%	86%

- Standard approach to "adjust" for differences is regression, $y = \beta_0 + \beta_1 \text{treat} + \beta_2 x_1 + \beta_3 x_2 + \epsilon$
- Estimate of β_1 would be reported as the treatment effect

Standard Regression Approach Finds an Effect When None Exists

- However, I generated the outcome so that there is no treatment effect
  $y = 0 + 0 \times \text{treat} + 1x_1 + 1x_2 + 4x_1x_2 + \text{N}(0,1)$
- Standard practice, which would fail to include the interaction, finds an effect when none exists

	Estimate	Std. Error	p value
(Intercept)	0.086	0.349	0.807
treat	-0.791	0.293	**0.008**
x1	2.654	0.260	0.000
x2	2.955	0.242	0.000

When Treatment Is Independent of X Conclusions Are Insensitive to Model

	Treatment	Control
y	2.74	2.73
x_1	60%	62%
x_2	61%	57%

	Estimate	Std. Error	p value
(Intercept)	-1.430	0.227	<0.001
treat	-0.051	0.202	**0.801**
x1	3.588	0.208	<0.001
x2	3.386	0.206	<0.001

Even No Adjustment for X Gives a Good Treatment Effect Estimate

	Treatment	Control
y	2.74	2.73
x_1	60%	62%
x_2	61%	57%

	Estimate	Std. Error	p value
(Intercept)	2.725	0.279	<0.001
treat	0.012	0.395	**0.975**

Regression Can Get the Right Answer… but Is Sensitive to Misspecification

- Correct treatment effect depends on including a critical interaction term
- With a large number of features this becomes hard
- Regression diagnostics are inadequate

	Estimate	Std. Error	p value
(Intercept)	0.149	0.264	0.573
treat	-0.103	0.227	**0.651**
x1	0.956	0.264	<0.001
x2	0.908	0.230	<0.001
x1*x2	3.865	0.336	<0.001

Idea: Reweight the Control Cases to Be Similar to the Treatment Cases

(x_1, x_2)	Treatment	Control	Needed weight	
0, 0	48	3	48/3=	16.0
1, 0	22	11	22/11=	2.0
0, 1	25	7	25/7=	3.57
1, 1	5	79	5/79=	0.06

Idea: Reweight the Control Cases to Be Similar to the Treatment Cases

y	Treatment	x_1	x_2	Weight
0.89	1	1	0	1
0.97	1	1	1	1
	…			
-0.01	0	0	0	16.00
0.55	0	1	0	2.00
-0.15	0	0	1	3.57
6.4	0	1	1	0.06
	…			

After Weighting the Groups Look Alike

	Treatment	Weighted Control	Unweighted Control
y	0.90	0.83	5.02
x_1	27%	27%	90%
x_2	30%	30%	86%

$$\bar{y}_1 = \frac{\sum_{t_i=1} y_i}{n_1}$$

$$\bar{y}_0 = \frac{\sum_{t_i=0} w_i y_i}{\sum_{t_i=0} w_i}$$

After Weighting the Groups Look Alike

	Treatment	Weighted Control	Unweighted Control
y	0.90	0.83	5.02
x_1	27%	27%	90%
x_2	30%	30%	86%

	Estimate	Std. Error	p value
(Intercept)	0.826	0.230	<0.001
treat	0.072	0.287	**0.803**

After Weighting the Groups Look Alike

	Treatment	Weighted Control	Unweighted Control
y	0.90	0.83	5.02
x_1	27%	27%	90%
x_2	30%	30%	86%

	Estimate	Std. Error	p value
(Intercept)	-0.149	0.259	0.565
treat	0.072	0.248	**0.772**
x1	1.764	0.284	<0.001
x2	1.661	0.311	<0.001

How to Calculate the Right Weights

$$\text{weight}(\mathbf{x}) = \frac{\text{number treated with } \mathbf{x}}{\text{number controls with } \mathbf{x}}$$

$$= \frac{n_1(\mathbf{x})}{n(\mathbf{x}) - n_1(\mathbf{x})}$$

$$= \frac{n_1(\mathbf{x})/n(\mathbf{x})}{n(\mathbf{x})/n(\mathbf{x}) - n_1(\mathbf{x})/n(\mathbf{x})}$$

$$= \frac{P(\text{treat} = 1|\mathbf{x})}{1 - P(\text{treat} = 1|\mathbf{x})}$$

x=(0,0)

weight=$\frac{48}{3}$

The Propensity Score

Propensity Score Weighting Is the Same as Reweighting Samples

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

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

$$= \frac{f(t = 1|\mathbf{x})f(\mathbf{x})f(t = 0)}{f(t = 0|\mathbf{x})f(\mathbf{x})f(t = 1)}$$

$$\propto \frac{f(t = 1|\mathbf{x})}{1 - f(t = 1|\mathbf{x})}$$

Using the Propensity Score Produces the Intuitive Weight We Derived Earlier

(x_1, x_2)	Treatment	Control	P(treat|x)	p/(1-p)
0, 0	48	3	48/51	16.0
1, 0	22	11	22/33	2.0
0, 1	25	7	25/32	3.57
1, 1	5	79	5/84	0.06

Regression Models Can Correctly Estimate Treatment Effects

- If the structure of the regression model is correct, then it too will correctly estimate the effect

$$E\big(Y(t)\big) = \beta_0 + \gamma t + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots$$

- Regression is not robust
  - Sensitive to missing interaction effects
  - Sensitive to missing non-linear relationship
  - Except in simple cases, you never know if your model is adequate

Propensity Scoring Just Introduces a New Problem

- Now we need to estimate $p(\mathbf{x})$
- Logistic regression is the common approach,
  $$\log \frac{p(\mathbf{x})}{1 - p(\mathbf{x})} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots$$
  but the same issues persist
  - Interactions
  - Transformations of features
  - Numerous features
  - Highly correlated features
  - Missing data
- Problems are easy to diagnose

Earlier Example of Propensity Score Cheated

- Creating four different propensity score weights for (0,0), (0,1), (1,0), and (1,1) is identical to fitting a logistic regression model with an interaction

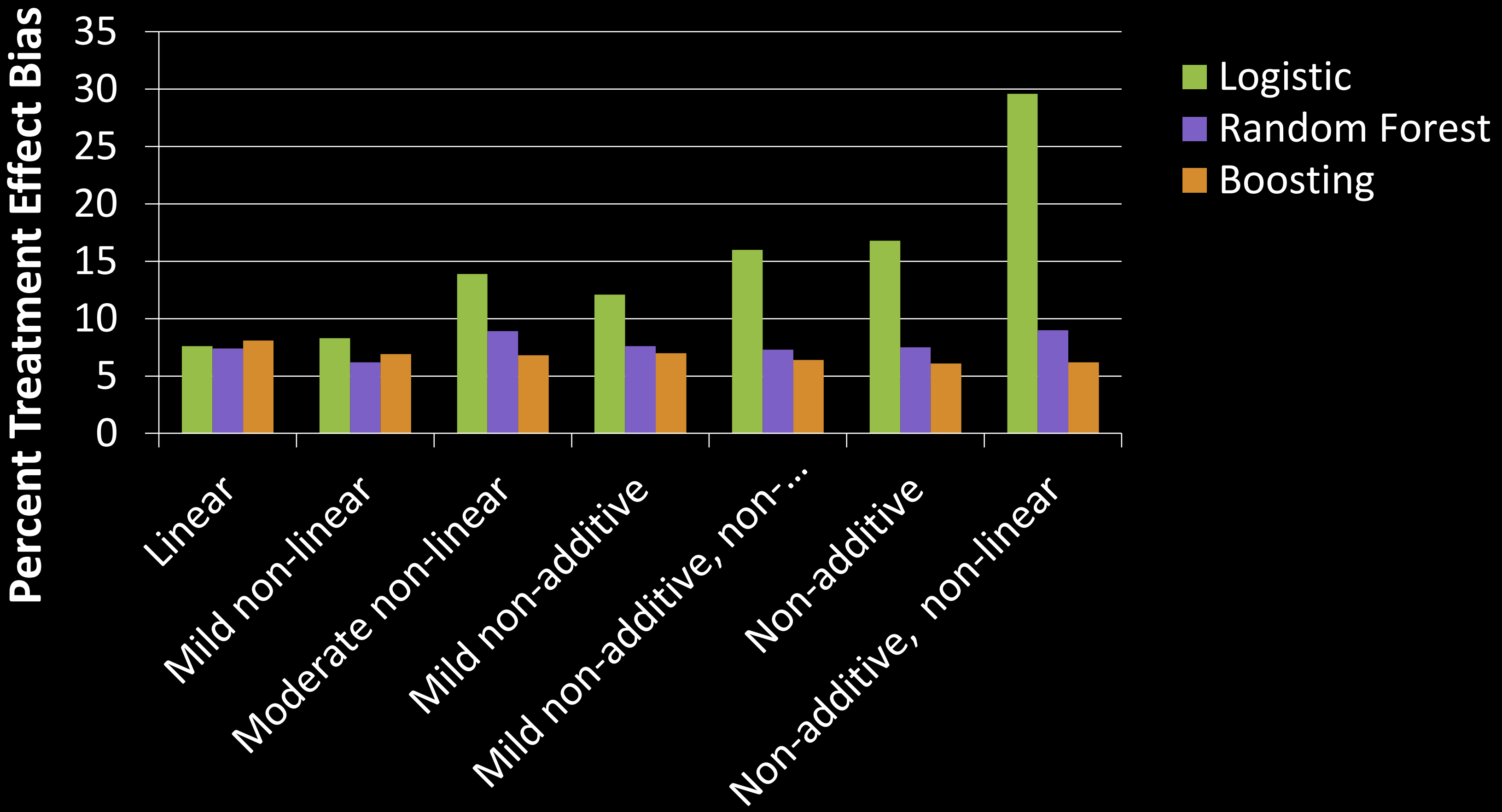
	Treatment	**Weighted Control (x_1*x_2)**		**Unweighted Control**
y	0.90	0.83		5.02
x_1	27%	27%		90%
x_2	30%	30%		86%

Earlier Example of Propensity Score Cheated

- Standard propensity score does not perform well
- Easy to assess that it did not work

	Treatment	Weighted Control (x_1*x_2)	Weighted Control (x_1+x_2)	Unweighted Control
y	0.90	0.83	0.60	5.02
x_1	27%	27%	14%	90%
x_2	30%	30%	14%	86%

Recommend Boosting for Estimating Propensity Scores



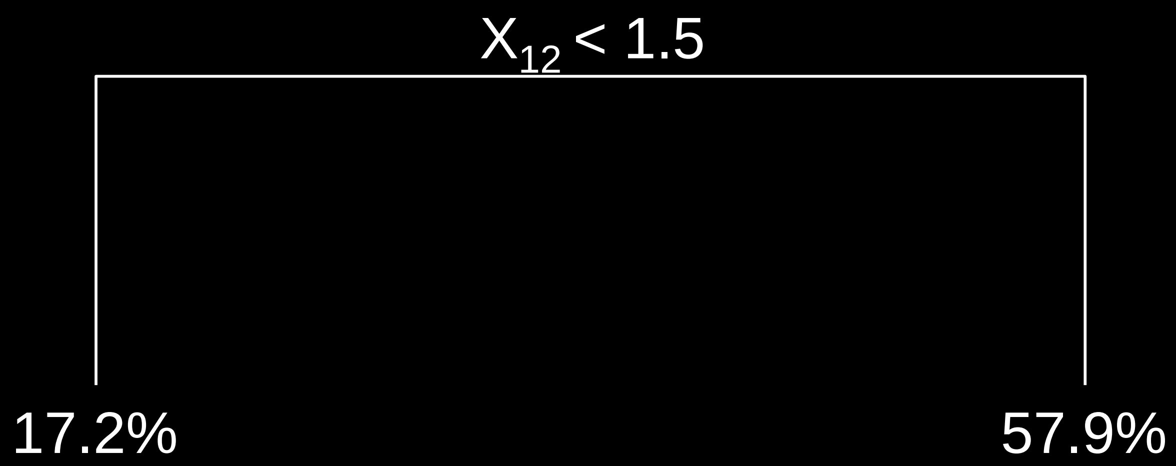
Results from Lee, Lessler, Stuart (2010)

Outline

- Natural experiments
- Propensity score/doubly robust methods
  - Race bias in post-stop outcomes
    - *Exercise*
  - Performance benchmarking officers and communities
- Additional topics in criminology and statistics

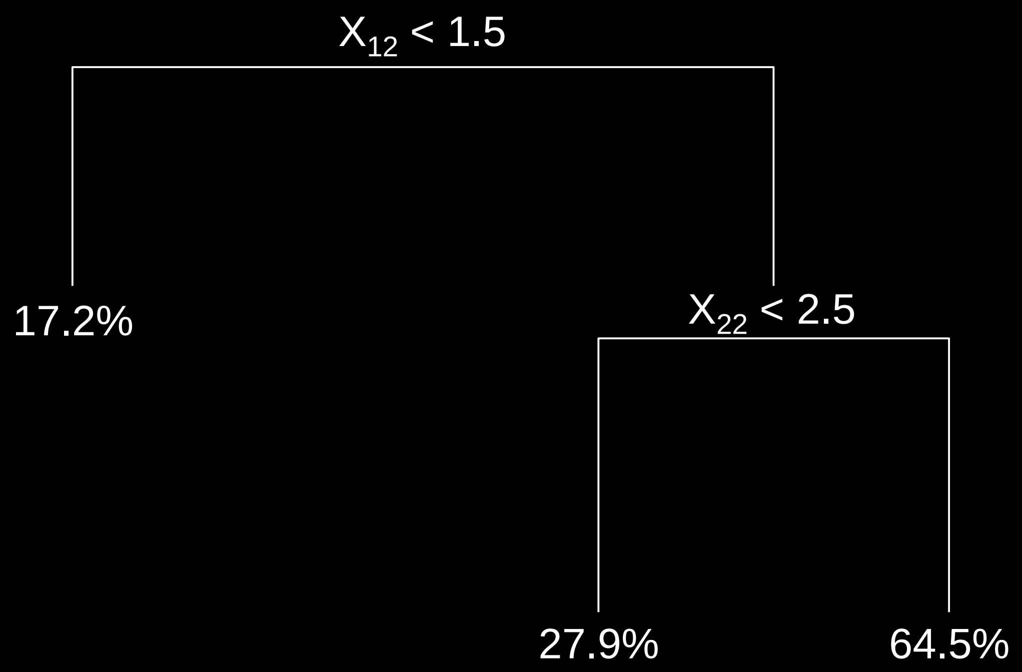
1. Search for a Good Split Point

- Find a variable and a split point that best estimates the probability of treatment assignment



2. Recursively Search for Improvements

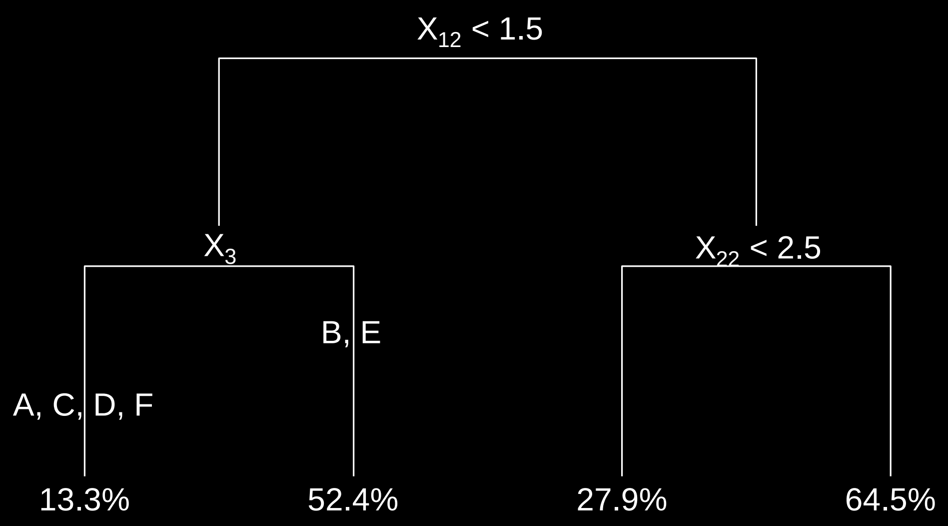
- Find another variable and split point that further refine the two groups



- Partitioning the right node offered the greatest improvement in predictive performance

3. Select a Stopping Point

- Controls the complexity of the interaction



- Stop when there is no further increase in predictive performance, too few observations, or when there is sufficient complexity

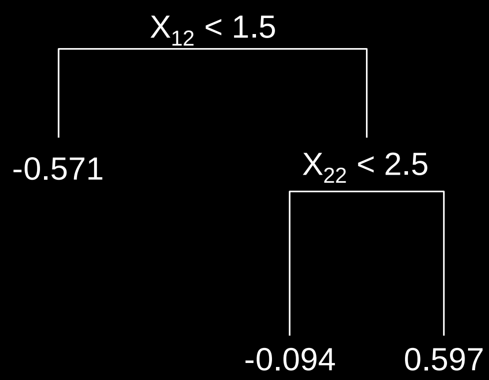
Regression Trees Have Pros and Cons

- Produce poor propensity scores
- Can be a building block for flexible estimation of propensity scores
- Handle continuous, ordinal, and categorical variables as well as missing data

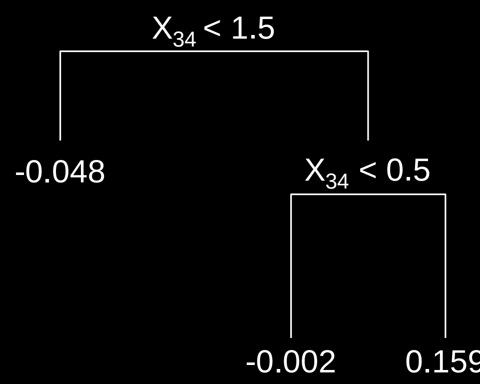
Boosting Replaces Linear Covariates with Tree Structured Terms

- $\log\left(\frac{p(\mathbf{x})}{1-p(\mathbf{x})}\right) = \beta_0 + \beta_1 \text{tree}_1(\mathbf{x}) + \beta_2 \text{tree}_2(\mathbf{x}) + \cdots$
  $+\beta_{5,000}\text{tree}_{5,000}(\mathbf{x})$

-0.452 +



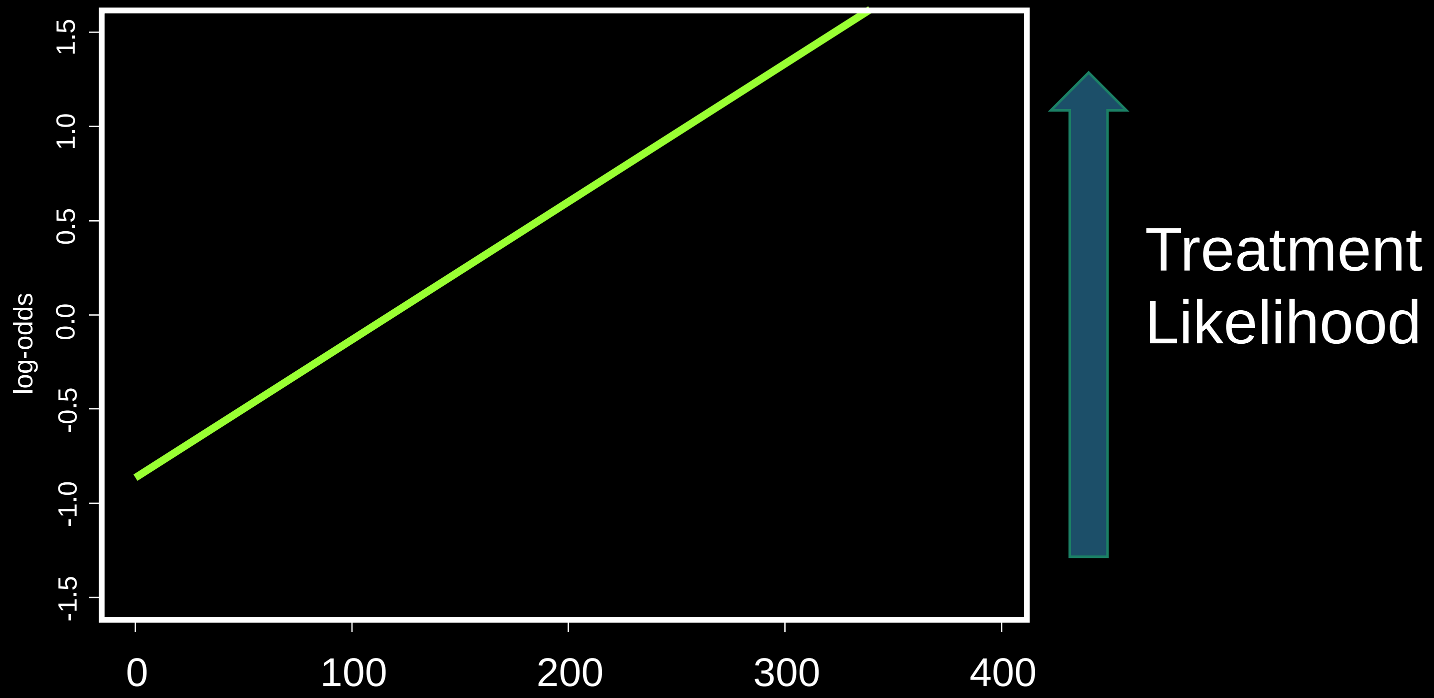
+



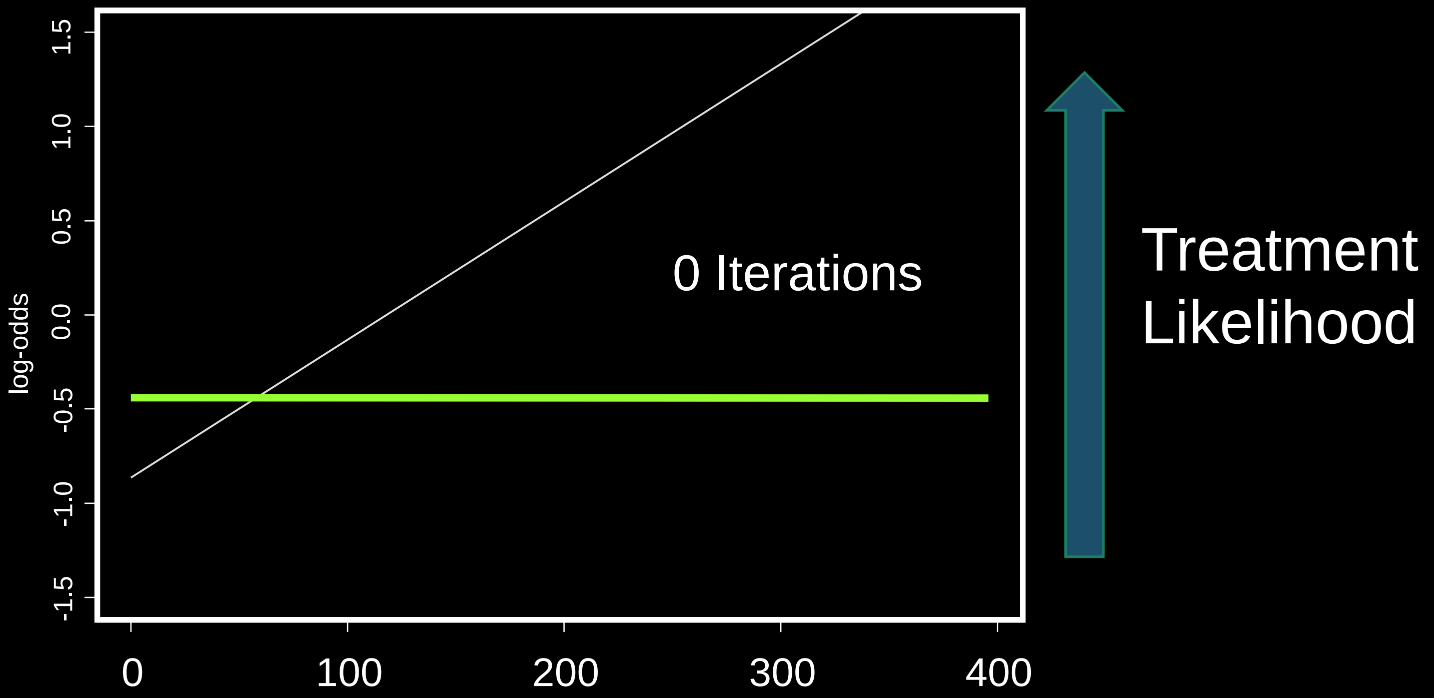
+ …

- Subject to a constraint that $\sum_{j=1}^{5,000} |\beta_j| \leq s$

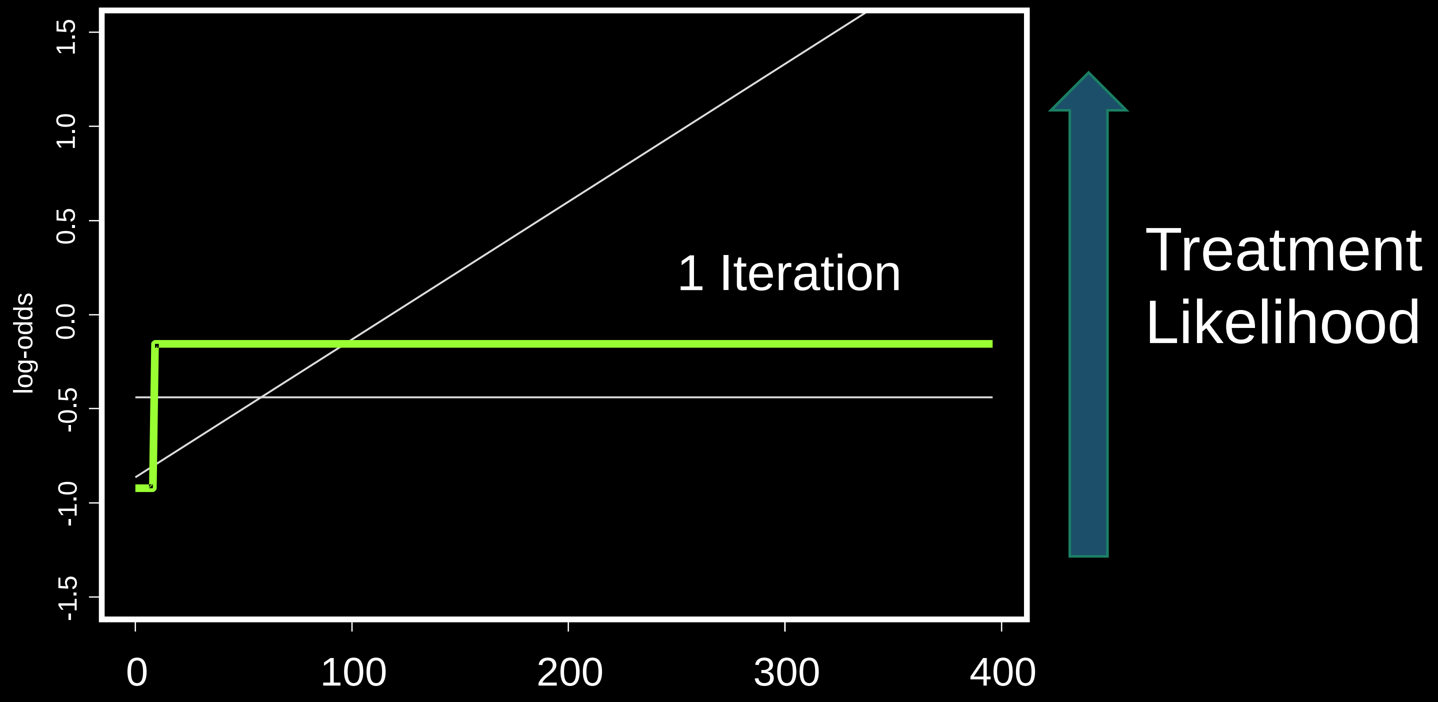
Logistic Regression



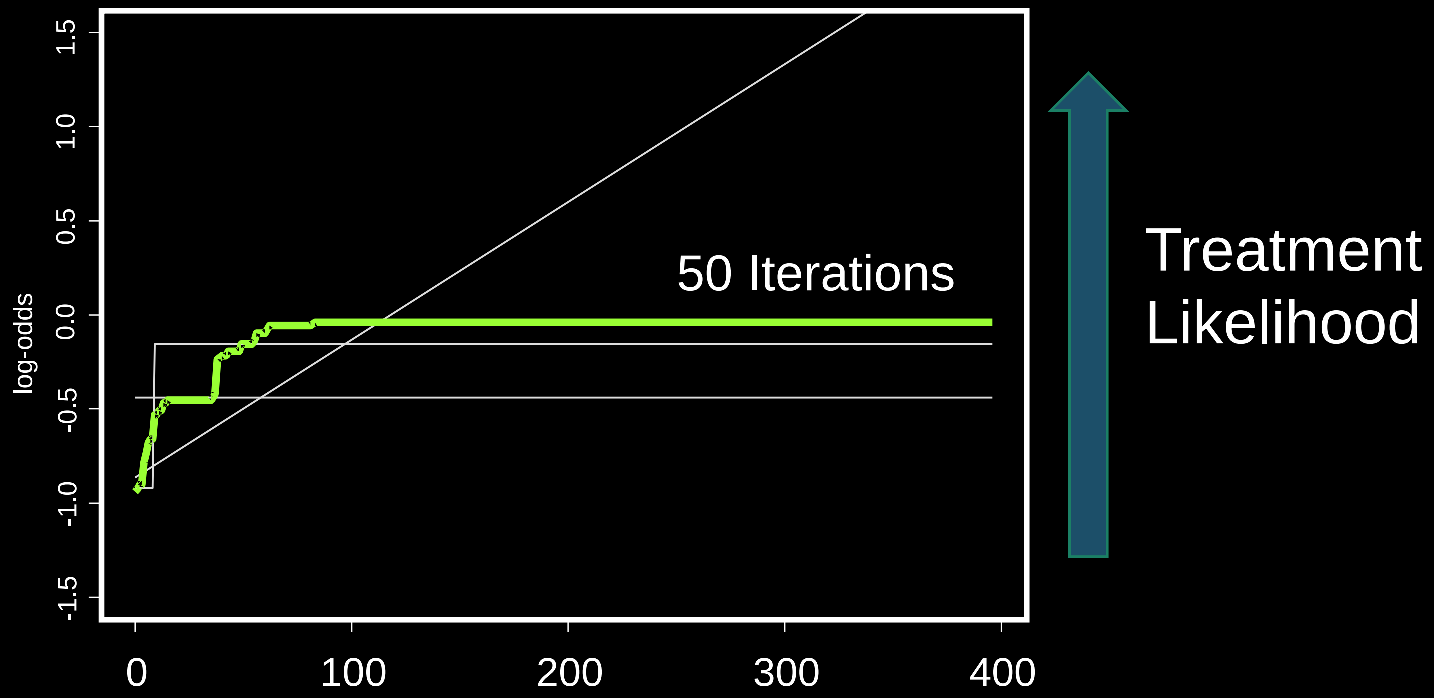
Boosted Propensity Score Estimation



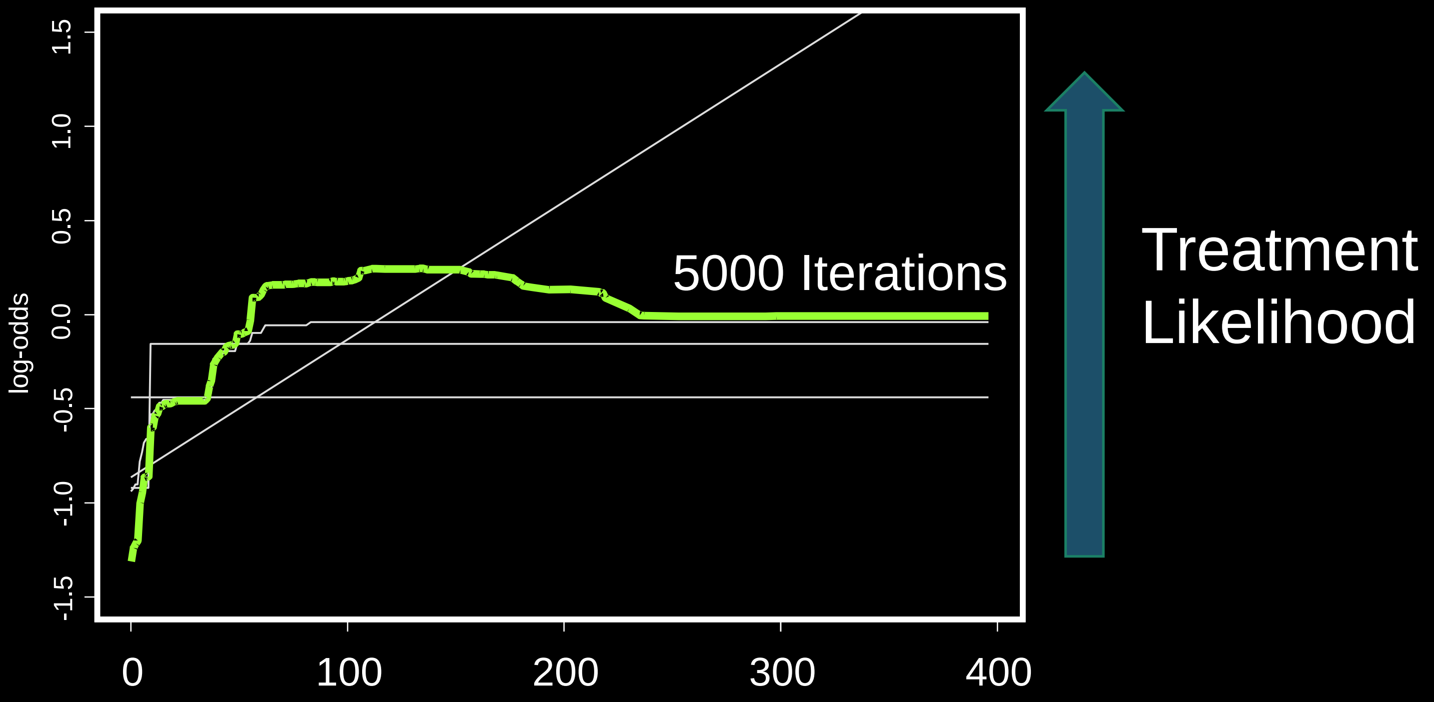
Boosted Propensity Score Estimation



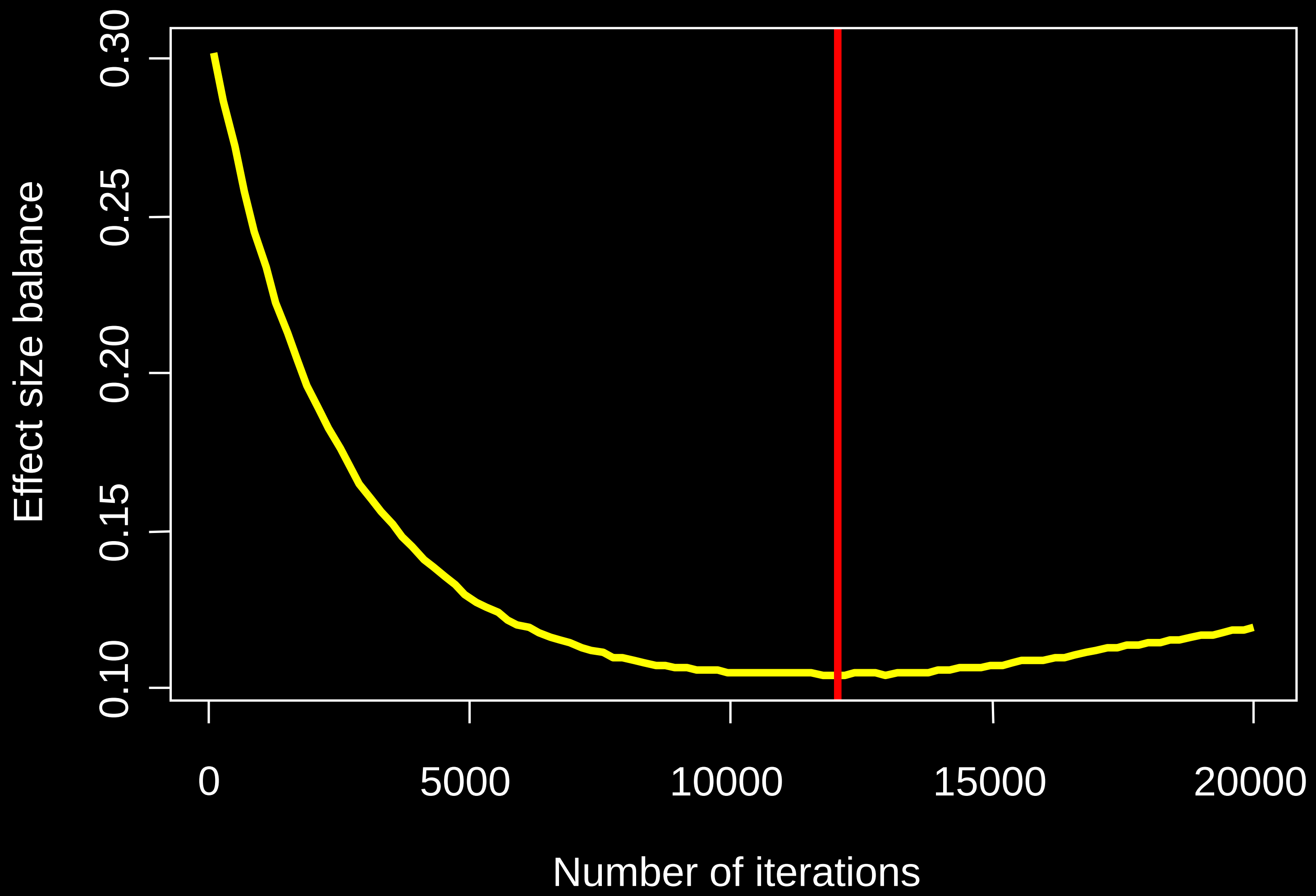
Boosted Propensity Score Estimation



Boosted Propensity Score Estimation



Continue Until the Control Group Looks the Most Like the Treatment Group



Numerous Advantages in Boosting, Computation Time Only Downside

- Excellent estimation of $p(\mathbf{x})$
- Balances the **x**s with little effort
- The resulting model handles continuous, nominal, ordinal, and missing **x**s
- Invariant to 1-to-1 transformations of the **x**s
- Model higher interaction terms with more complex regression trees

Return to Exercise

Outline

- Natural experiments
- Propensity score/doubly robust methods
  - Race bias in post-stop outcomes
    - *Exercise*
  - Performance benchmarking officers and communities
- Additional topics in criminology and statistics

Ridgeway & MacDonald (2014). "A Method for Internal Benchmarking of Criminal Justice System Performance," *Crime & Delinquency* 60(1):145-162

Ridgeway & MacDonald (2009). "Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops," *Journal of the American Statistical Association* 104(486):661–668

Is an Officer Who Stops 86% Black Pedestrians Unusual?

Stop Characteristic	Example Officer (%) n = 392	
% black pedestrians stopped	86%	

- Combine three statistical techniques to answer this question
  - Propensity score weighting
  - Doubly robust estimation
  - False discovery rate

G. Ridgeway and J.M. MacDonald (2009). “Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops.” JASA 104:661–668

We Know a Lot About the Environment of this Officer's Stops

Stop Characteristic		Example Officer (%) n = 392	
% black pedestrians stopped		86%	
Month	January	3	
	February	4	
	March	8	
Day of the week	Monday	13	
	Tuesday	11	
	Wednesday	14	
Time of day	(4-6 p.m.]	9	
	(6-8 p.m.]	8	
	(8-10 p.m.]	23	
	(10 p.m. -12 a.m.]	17	
Patrol borough	Brooklyn North	100	
Precinct	B	98	
	C	1	
Outside		96	
In uniform	Yes	99	
Radio run	Yes	1	

We Also Know the Exact Location of This Officer's Stops

Example Officer

Idea: Reweight Stops Made By Other Officers to Resemble This Officer's Stops

Example Officer

- Align their distributions

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

- Solving for $w(\mathbf{x})$ yields the propensity score weight

$$w(\mathbf{x}) \propto \frac{P(t = 1|\mathbf{x})}{1 - P(t = 1|\mathbf{x})}$$

- Estimate $P(t = 1|\mathbf{x})$ using boosted logistic regression as implemented in `gbm`

Reweighting Aligns the Distribution of Stop Locations

Example Officer

Matched Stops

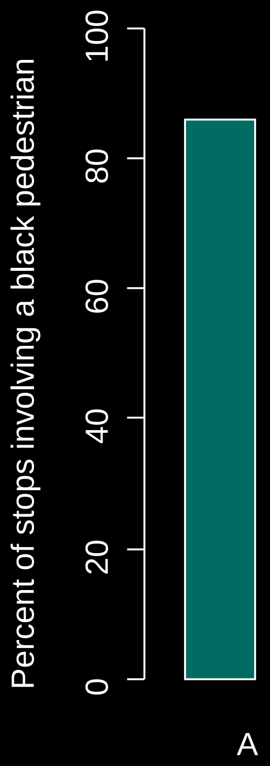
Reweighting Also Aligns the Distribution of All Other Stop Features

Stop Characteristic		Example Officer (%) n = 392	Internal Benchmark (%) ESS = 3,676
	% black pedestrians stopped	**86%**	
Month	January	3	3
	February	4	4
	March	8	9
Day of the week	Monday	13	13
	Tuesday	11	10
	Wednesday	14	15
Time of day	(4-6 p.m.]	9	10
	(6-8 p.m.]	8	8
	(8-10 p.m.]	23	23
	(10 p.m. -12 a.m.]	17	17
Patrol borough	Brooklyn North	100	100
Precinct	B	98	98
	C	1	1
Outside		96	94
In uniform	Yes	99	97
Radio run	Yes	1	3

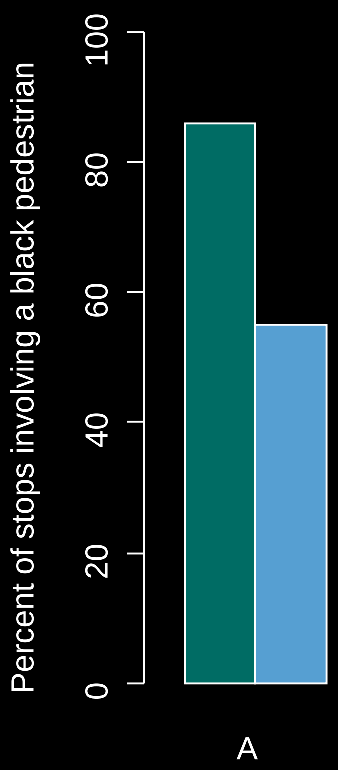
Colleagues at the Same Time, Place, and Context Stop 55% Black Pedestrians

Stop Characteristic		Example Officer (%) n = 392	Internal Benchmark (%) ESS = 3,676
% black pedestrians stopped		**86%**	**55%**
Month	January	3	3
	February	4	4
	March	8	9
Day of the week	Monday	13	13
	Tuesday	11	10
	Wednesday	14	15
Time of day	(4-6 p.m.]	9	10
	(6-8 p.m.]	8	8
	(8-10 p.m.]	23	23
	(10 p.m. -12 a.m.]	17	17
Patrol borough	Brooklyn North	100	100
Precinct	B	98	98
	C	1	1
Outside		96	94
In uniform	Yes	99	97
Radio run	Yes	1	3

86% of the Officer's Stops Were Black…



…Compared with 55% for the Benchmark



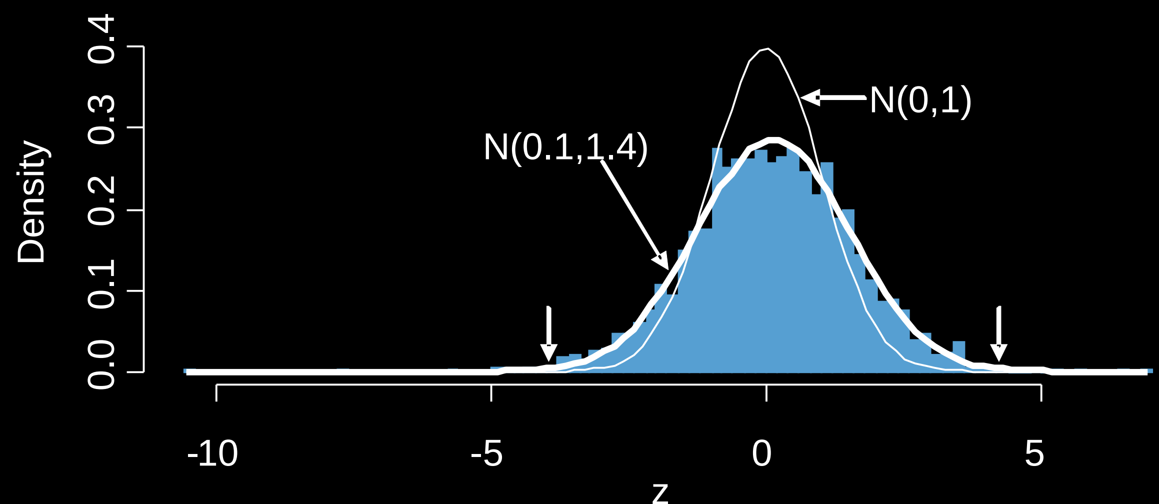
- Doubly robust benchmark estimate obtainable from weighted logistic regression

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} w_i \left(y_i s(t_i, \mathbf{x}_i | \boldsymbol{\beta}) - \log\left(1 + e^{s(t_i, \mathbf{x}_i | \boldsymbol{\beta})}\right)\right)$$

- Disparity computed as

$$\hat{\theta}_A^{DR} = \sum_{i=1}^{n} t_i \left(\frac{1}{1 + \exp\left(-s\left(1, \mathbf{x}_i | \hat{\boldsymbol{\beta}}\right)\right)} - \frac{1}{1 + \exp\left(-s\left(0, \mathbf{x}_i | \hat{\boldsymbol{\beta}}\right)\right)} \right)$$

Repeat for Nearly 3,000 NYPD Officers Actively Involved in Stops

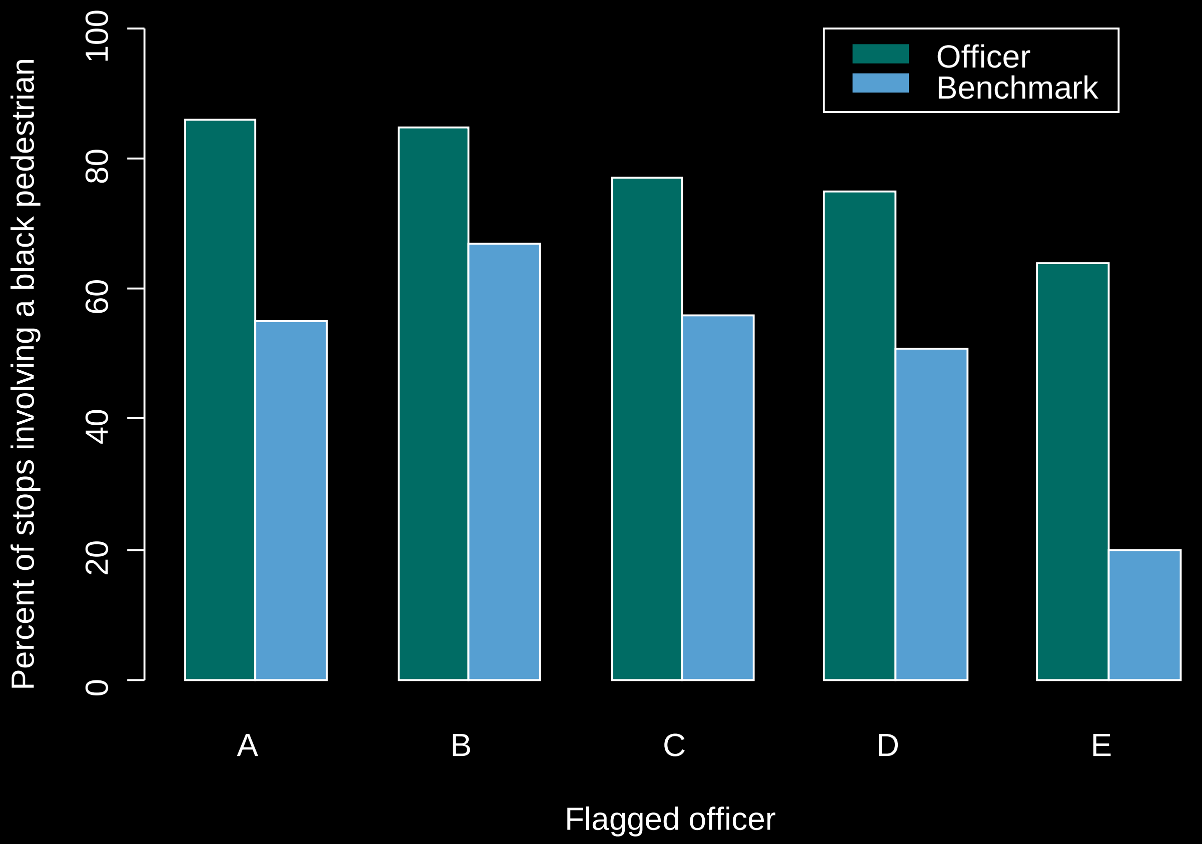


- $P(\text{problem}|z) = 1 - \frac{f(z|\text{no problem})f(\text{no problem})}{f(z)}$

  $$\geq 1 - \frac{f_0(z)}{f(z)}$$

- Right tail consists of 5 officers with "problem officer" probabilities in excess of 50%
- Standard cutoff of z > 2.0 flags 242 officers, 90% of which have fdr estimated to be greater than 0.999

Analysis in NYPD Flagged Five Officers



When We Return…

- Natural experiments
- Propensity score/doubly robust methods
- Additional topics in criminology and statistics
  - Which police officers are most likely to shoot?
  - Place-based interventions
    - Gang injunctions
    - Homeless shelters

Outline

- Natural experiments
- Propensity score/doubly robust methods
- Additional topics in criminology and statistics
  - Which police officers are most likely to shoot?
  - Place-based interventions
    - Gang injunctions
    - Homeless shelters

Ridgeway (2016). "Officer Risk Factors Associated with Police Shootings: A Matched Case-Control Study," *Statistics and Public Policy*

Police Use of Lethal Force Sparks Unrest

- 2001 Cincinnati PD shooting of Timothy Thomas resulted in 4 days of riots and $3.6M in damage

- 2006 NYPD shooting of Sean Bell, 50 shots fired. Officers found not guilty at trial, but fired or resigned

- 2014 Chicago PD shooting of Laquan McDonald. 16 bullets fired by one officer, no other officer fired

McElvain and Kposowa (2008) Compared Shooters to Non-Shooters

- Riverside County Sheriff Department
  - 186 shooting incidents involving 314 deputies
  - Control group consisted of 334 deputies with no involvement in shooting incidents
  - Data for shooters collected at time of shooting, controls collected in 2004
- Shooters were more likely to be male, Hispanic, no college, younger, and in lower ranks
- Unmeasured confounding is a major concern in such a study design

Fyfe (1989) states that "there is virtually no empirical support for assertions that individual officer characteristics are measurably related to any type of performance in office"

NYPD Sought a Comprehensive Review of Firearm Practices

- Prompted by controversy surrounding an officer-involved shooting, NYPD Police Commissioner sought a review of:
  - Initial firearms training provided to new recruits
  - In-service firearms training
  - Firearms Discharge Review Board functions and processes
  - The phenomenon of reflexive shooting

"The characteristics of officers involved in discharge incidents will be examined for patterns in training, experience, supervision, and other factors that may help predict, and thus reduce, firearms discharges generally and inappropriate discharges in particular"

Assessing Officer Risk Factors Requires Controlled Comparison

- Officers that discharge their weapons often look different from other officers in obvious ways, such as
  - In the field
  - In particular neighborhoods
  - Conducting higher risk operations
  - Not at a desk
- Idea: Compare shooting officer to other non-shooting officers on the scene
  - Does not judge shooting justification
  - But if there is a consistent pattern it could inform training or assignments

Each Shooting Is an Experiment

1. Multiple officers on the scene
2. Each officer has a latent risk of shooting
3. Before the shooting, each officer on the scene could have been the shooter
4. Test whether there are officer features that affect the risk of shooting

Odds of Shooting Depend on Environment and Officer Features

$$\textbf{odds of shooting} = \mathbf{1.1} \times \mathbf{1.6} \times \mathbf{0.9}$$

- Officer is in a high risk environment

Odds of Shooting Depend on Environment and Officer Features

$$\textbf{odds of shooting} = \mathbf{1.1} \times \mathbf{1.6} \times \mathbf{0.9}$$

- Officer is in a high risk environment
- Officer has many negative marks in file

Odds of Shooting Depend on Environment and Officer Features

$$\textbf{odds of shooting} = \mathbf{1.1} \times \mathbf{1.6} \times \mathbf{0.9}$$

- Officer is in a high risk environment
- Officer has many negative marks in file
- Officer joined NYPD at age 30

Odds of Shooting Depend on Environment and Officer Features

$$\textbf{odds of shooting} = \mathbf{1.1} \times \mathbf{1.6} \times \mathbf{0.9} = \mathbf{1.58}$$

- Officer is in a high risk environment
- Officer has many negative marks in file
- Officer joined NYPD at age 30

Very difficult to collect enough data on environment and monitor police in all scenarios to estimate these risk factors

One of Two Officers Shoot, We Can Guess Who Shot

$$\frac{P(A\ \text{shoot})}{1-P(A\ \text{shoot})} = 1.1 \times 1.6 \times 0.9$$

$$\frac{P(B\ \text{shoot})}{1-P(B\ \text{shoot})} = 1.1 \times 0.5 \times 1.1$$

$$P(A\ \text{shoots} | A\ \text{or}\ B\ \text{shoots}) = \frac{1.1 \times 1.6 \times 0.9}{1.1 \times 1.6 \times 0.9 + 1.1 \times 0.5 \times 1.1}$$

Low Risk Environment? Environment Cancels Out

$$\frac{P(A\ \text{shoot})}{1 - P(A\ \text{shoot})} = 0.2 \times 1.6 \times 0.9$$

$$\frac{P(B\ \text{shoot})}{1 - P(B\ \text{shoot})} = 0.2 \times 0.5 \times 1.1$$

$$P(A\ \text{shoots} | A\ \text{or}\ B\ \text{shoots}) = \frac{0.2 \times 1.6 \times 0.9}{0.2 \times 1.6 \times 0.9 + 0.2 \times 0.5 \times 1.1}$$

High Risk Environment?
Environment Cancels Out

$$\frac{P(A\ \text{shoot})}{1 - P(A\ \text{shoot})} = 5.0 \times 1.6 \times 0.9$$

$$\frac{P(B\ \text{shoot})}{1 - P(B\ \text{shoot})} = 5.0 \times 0.5 \times 1.1$$

$$P(A\ \text{shoots} | A\ \text{or}\ B\ \text{shoots}) = \frac{5.0 \times 1.6 \times 0.9}{5.0 \times 1.6 \times 0.9 + 5.0 \times 0.5 \times 1.1}$$

One of Two Officers Shoot, We Can Guess Who Shot

$$\frac{P(A\ \text{shoot})}{1 - P(A\ \text{shoot})} = 1.1 \times 1.6 \times 0.9$$

$$\frac{P(B\ \text{shoot})}{1 - P(B\ \text{shoot})} = 1.1 \times 0.5 \times 1.1$$

$$\frac{1.6 \times 0.9}{1.6 \times 0.9 + 0.5 \times 1.1} = 0.72$$

Chance that Officer A shot depends only on her features
We do not need to measure environmental features

I Solve the Inverse Problem, Observe Who Shoots, Infer Parameters

$$\frac{P(A\ \text{shoot})}{1 - P(A\ \text{shoot})} = \cancel{1.1} \times \beta_1 \times \beta_2$$

$$\frac{P(B\ \text{shoot})}{1 - P(B\ \text{shoot})} = \cancel{1.1} \times \beta_3 \times \beta_4$$

Shooter

Nonshooter

Learn the Factors Affecting the Probability of Shooting

$$\log \frac{P(S = 1|\mathbf{x}, \mathbf{z})}{1 - P(S = 1|\mathbf{x}, \mathbf{z})} = h(\mathrm{z}) + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_d x_d$$

- *S* indicates that the officer shoots
- **x** are the officer's features
- **z** are the features of a particular scenario (kinds of suspects involved, location, and lighting)

Collected data do not quite match this framework

Utilized Data on a Review of Three Years of OIS Records

- Gathered data on all officer-involved shootings adjudicated in 2004, 2005, and 2006
- For each shooting I recorded
  - department ID numbers for shooters in the incident
  - department ID numbers for non-shooting officers that were witnesses or in the immediate vicinity of the shooting
- 106 incidents involving 150 shooting officers and 141 non-shooting officers
- Collected data on age, experience, education, training, and past performance

Consider the Likelihood of a Shooting Involving Two Officers

$$P(S_A = 1, S_B = 0 | S_A + S_B = 1, \mathbf{x}_A, \mathbf{x}_B, \mathbf{z}) =$$

$$\frac{P(S_A + S_B = 1 | S_A = 1, S_B = 0, \mathbf{x}_A, \mathbf{x}_B, \mathbf{z}) P(S_A = 1, S_B = 0 | \mathbf{x}_A, \mathbf{x}_B, \mathbf{z})}{P(S_A + S_B = 1 | \mathbf{x}_A, \mathbf{x}_B, \mathbf{z})} =$$

$$\frac{P(S_A = 1 | \mathbf{x}_A, \mathbf{z}) P(S_B = 0 | \mathbf{x}_B, \mathbf{z})}{P(S_A = 1 | \mathbf{x}_A, \mathbf{z}) P(S_B = 0 | \mathbf{x}_B, \mathbf{z}) + P(S_A = 0 | \mathbf{x}_A, \mathbf{z}) P(S_B = 1 | \mathbf{x}_B, \mathbf{z})}$$

Substituting Simplifies the Model

$$\frac{P(S_A = 1|\mathbf{x}_A, \mathbf{z})P(S_B = 0|\mathbf{x}_B, \mathbf{z})}{P(S_A = 1|\mathbf{x}_A, \mathbf{z})P(S_B = 0|\mathbf{x}_B, \mathbf{z}) + P(S_A = 0|\mathbf{x}_A, \mathbf{z})P(S_B = 1|\mathbf{x}_B, \mathbf{z})} =$$

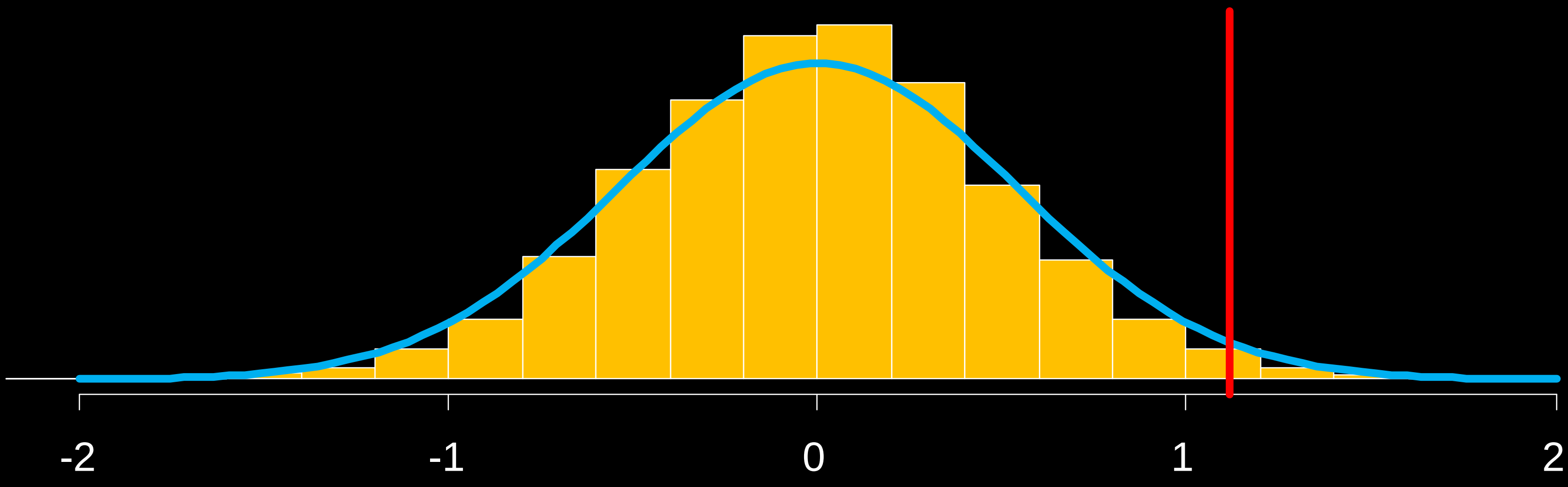
$$\frac{\frac{e^{h(\mathbf{z})+\beta' x_A}}{1+e^{h(\mathbf{z})+\beta' x_A}} \frac{1}{1+e^{h(\mathbf{z})+\beta' x_B}}}{\frac{e^{h(\mathbf{z})+\beta' x_A}}{1+e^{h(\mathbf{z})+\beta' x_A}} \frac{1}{1+e^{h(\mathbf{z})+\beta' x_B}} + \frac{1}{1+e^{h(\mathbf{z})+\beta' x_A}} \frac{e^{h(\mathbf{z})+\beta' x_B}}{1+e^{h(\mathbf{z})+\beta' x_B}}} =$$

$$\frac{e^{h(\mathbf{z})+\beta' x_A}}{e^{h(\mathbf{z})+\beta' x_A} + e^{h(\mathbf{z})+\beta' x_B}} =$$

$$\frac{e^{\beta' x_A}}{e^{\beta' x_A} + e^{\beta' x_B}}$$

Under the Null, the Shooting Labels Unrelated to Officer Features

- Randomly shuffle the shooting indicator *within each shooting* incident 10,000 times
- Refit conditional logistic regression model on each shuffled dataset
- Extract the coefficients



Who Is More Likely to Shoot?

Variable	Risk difference
Rank	
Police officer (reference)	
Detective	
Sergeant	
Lieutenant	
Captain	

Supervisors and Management Ranks Are Less Likely to Shoot

Variable	Risk difference
Rank	
Police officer (reference)	
Detective	No difference
Sergeant	-74%
Lieutenant	-95%
Captain	-96%

Who Is More Likely to Shoot?

Variable	Risk difference
Rank	
Police officer (reference)	
Detective	No difference
Sergeant	-74%
Lieutenant	-95%
Captain	-96%
Male	

Men and Women Equally Likely to Shoot

Variable	Risk difference
Rank	
Police officer (reference)	
Detective	No difference
Sergeant	-74%
Lieutenant	-95%
Captain	-96%
Male	No difference

Who Is More Likely to Shoot?

Variable	Risk difference
Rank	
Police officer (reference)	
Detective	No difference
Sergeant	-74%
Lieutenant	-95%
Captain	-96%
Male	No difference
Race	
White (reference)	
Black	
Hispanic	

Black Officers More Likely to Shoot

Variable	Risk difference
Rank	
Police officer (reference)	
Detective	No difference
Sergeant	-74%
Lieutenant	-95%
Captain	-96%
Male	No difference
Race	
White (reference)	
Black	+226%
Hispanic	No difference

Older Recruits Have a Sustained Lower Risk of Shooting

Variable	Risk difference
Rank	
Police officer (reference)	
Detective	No difference
Sergeant	-74%
Lieutenant	-95%
Captain	-96%
Male	No difference
Race	
White (reference)	
Black	+226%
Hispanic	No difference
Years at NYPD	No difference
Age when recruited	-11%
Education	No difference
Special assignment	No difference

What Kinds of Prior Activity Signal Increased Shooting Risk?

Variable	Risk difference
Average annual	
Evaluation score < 3.5	
Range score < 86	
Complaints > 0.6	
Medal count > 3.8	
CPI points > 3.1	
Gun arrests > 2.4	
Felony arrests > 9.3	
Misdemeanor arrests > 10.0	
Days of leave	

Rapid Accumulation of Negative Marks Signals Elevated Shooting Risk

Variable	Risk difference
Average annual	
Evaluation score < 3.5	
Range score < 86	
Complaints > 0.6	
Medal count > 3.8	
CPI points > 3.1	+212%
Gun arrests > 2.4	
Felony arrests > 9.3	
Misdemeanor arrests > 10.0	-80%
Days of leave	

8% of NYPD officers
15% of shooting scene officers

Central Personnel Index Assign Points to Problematic Incidents

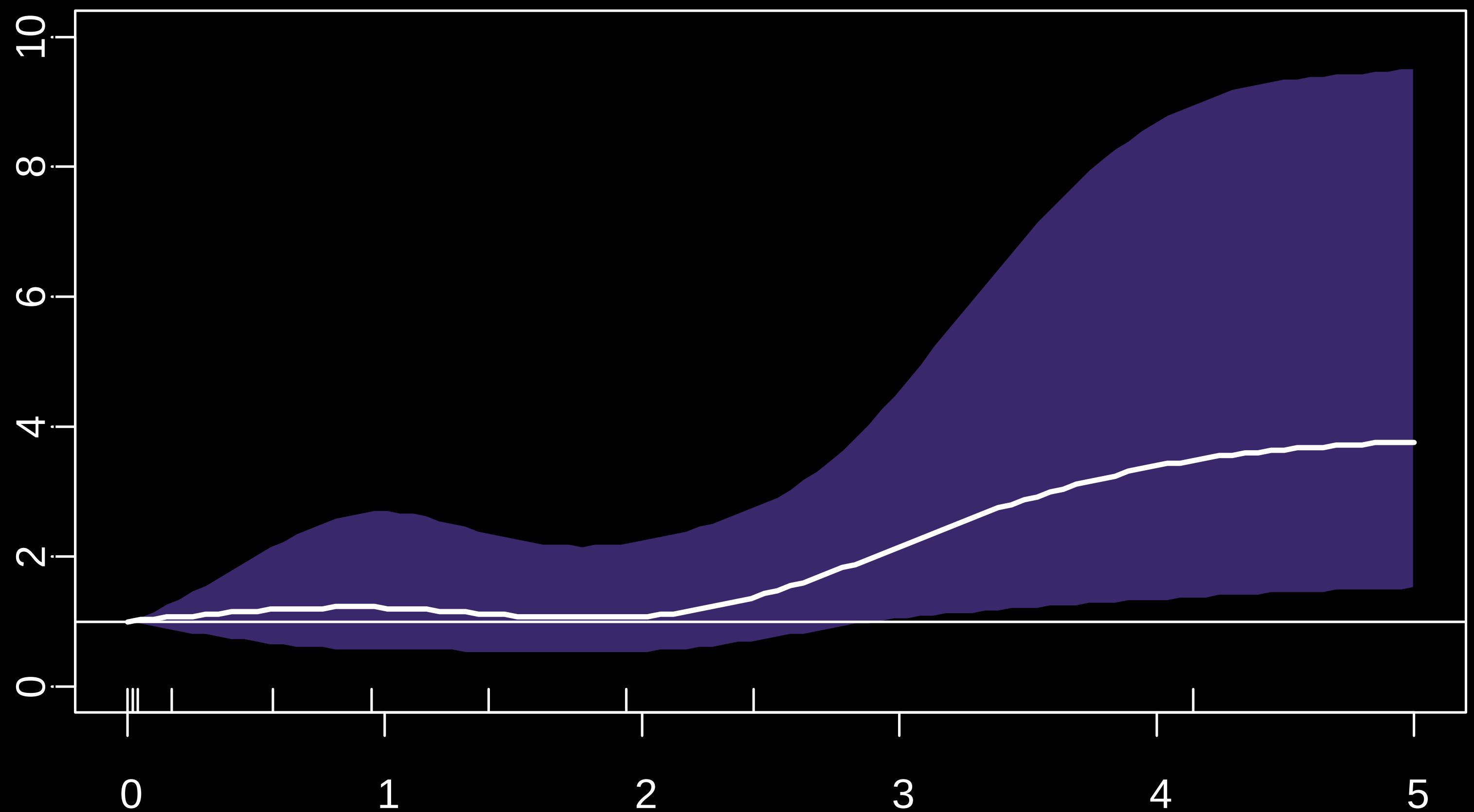
Event	Point value
Suspension	8
Loss of firearm	6
Negative evaluation - A	5
Fail to safeguard weapon	5
Chronic sick – B	4
Loss of shield	4
Negative evaluation – B	3
Chronic sick – A	2
Firearm discharge	1
Dept. auto accident	1

```
NEGATIVE EVALUAT. - B        10 MONTH EVAL - 3.0
DATE       : 04/30/2005      (1) LOW - BEHAV DIMENS
CONTROL #: 003
SERIAL   #: XXXX

FIREARMS DISCHARGE           NO VIOLATION
DATE       : 06/09/2006      NO CORRECTIVE ACTION
CONTROL #: 004
```

Exceeding 3.1 CPI/year Strongly Associated with Shooting Risk

Odds of being a shooting officer relative to officers with zero CPI points



ASW Mar 2017

"Active" Officer May Be Key Factor

Variable	Risk difference
Average annual	
Evaluation score < 3.5	No difference
Range score < 86	No difference
Complaints > 0.6	+107%
Medal count > 3.8	+128%
CPI points > 3.1	+212%
Gun arrests > 2.4	No difference
Felony arrests > 9.3	+115%
Misdemeanor arrests > 10.0	-80%
Days of leave	No difference

Outline

- Natural experiments
- Propensity score/doubly robust methods
- Additional topics in criminology and statistics
  - Which police officers are most likely to shoot?
  - Place-based interventions
    - Gang injunctions
    - Homeless shelters

Ridgeway, Moyer, MacDonald, Grogger (?). "Effect of Gang Injunctions on Crime: A Study of Los Angeles from 1988-2014"

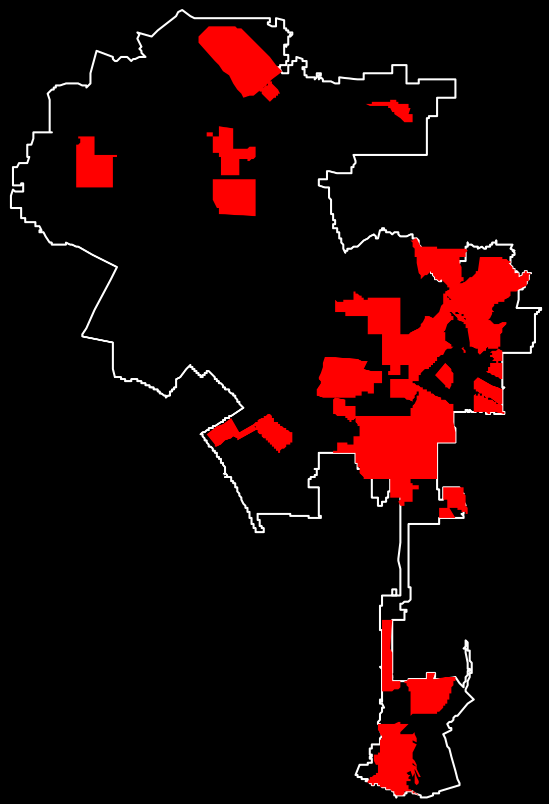
Faraji, Ridgeway, Wu (2017, under review). "Effect of Homeless Shelters on Crime: A Study of Vancouver, Canada from 2006-2015," *Journal of Experimental Criminology*

Civil Gang Injunctions (CGIs) are Neighborhood-Focused Interventions

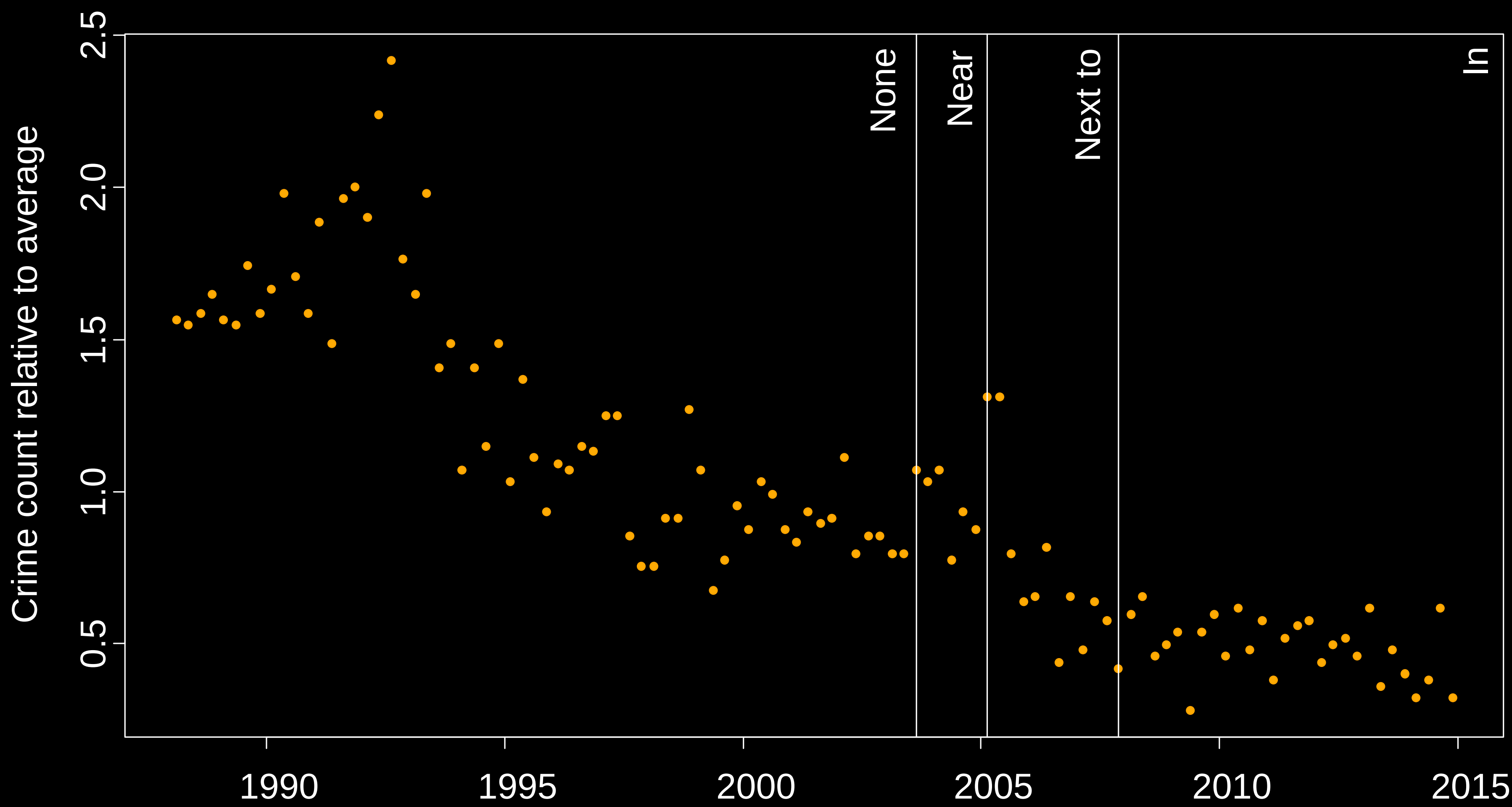
- Designed to interfere with routine behaviors of gang members within defined geographic areas ("safety zones")
- Civilly enjoin otherwise-legal activities (e.g., publically congregating with other gang members) as well as criminal offenses (e.g., drug trafficking)

Los Angeles Provides Useful Framework to Assess CGI Effects

- LA historically has experienced severe gang-related crime
- 48 CGIs currently in effect in LA; 3 earlier CGIs terminated
- Any effects begin when complaint is served
- Our analysis uses quarterly LAPD crime reports (1988-2014)
- 939 RDs over 108 quarters



RD1204 Transitions from No Safety Zone, to Near, Next to, in a Safety Zone

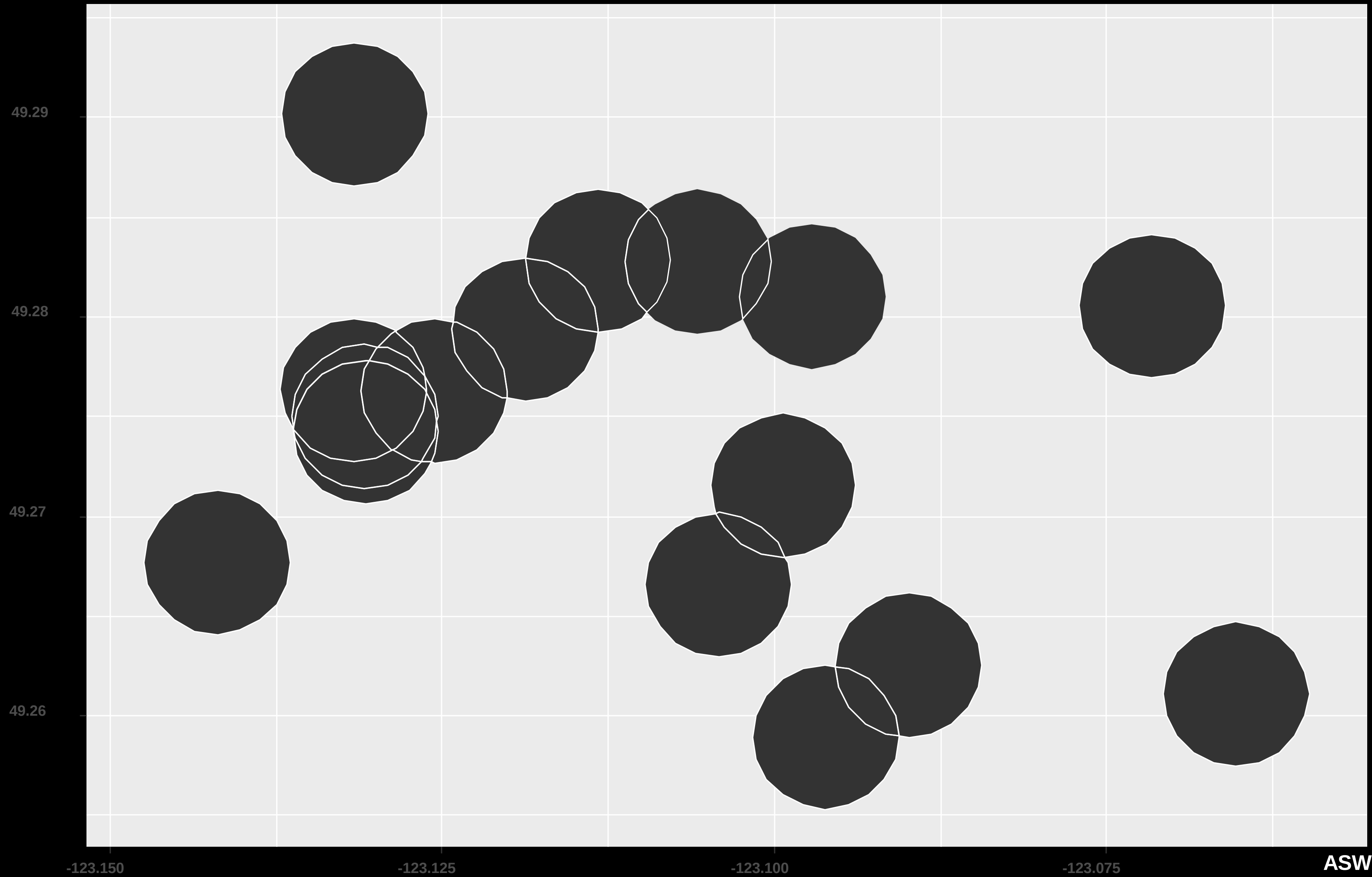


ASW Mar 2017

Vancouver Launched a Winter Emergency Shelter Program in 2009

Shelter	2009	2010	2011	2012	2013	2014	2015	2016
51B W Cordova Street	✓	✓	✓	✓	✓	✓	✓	✓
320 Hastings Street	✓	✓	✓	✓	✓	✓	✓	✓
201 Central Street	✓	✓	✓	✓	✓	✓	✓	✓
134 East Cordova Street	✓	✓	✓	✓	✓	✓	✓	✓
1442 Howe Street	✓		✓	✓				
1435 Granville Street	✓	✓						
1642 West 4th Avenue		✓	✓					
747 Cardero Street		✓	✓					
677 East Broadway Street		✓	✓					
1648 East 1st Avenue		✓	✓					✓
518 Richards Street				✓				
2950 Prince Edward Street				✓				
900 Pacific Street				✓	✓	✓	✓	✓
119 East Cordova Street				✓				✓
1210 Seymour Street					✓			
2610 Victoria Drive					✓			
21 East 5th Avenue					✓	✓		
862 Richards Street					✓	✓		
1647 East Pender Street							✓	

400m Buffers Around Each Shelter Generate 31 Distinct Regions



ASW Mar 2017

Stepped Wedge Design Aims to Detect Shifts in Crime Rates

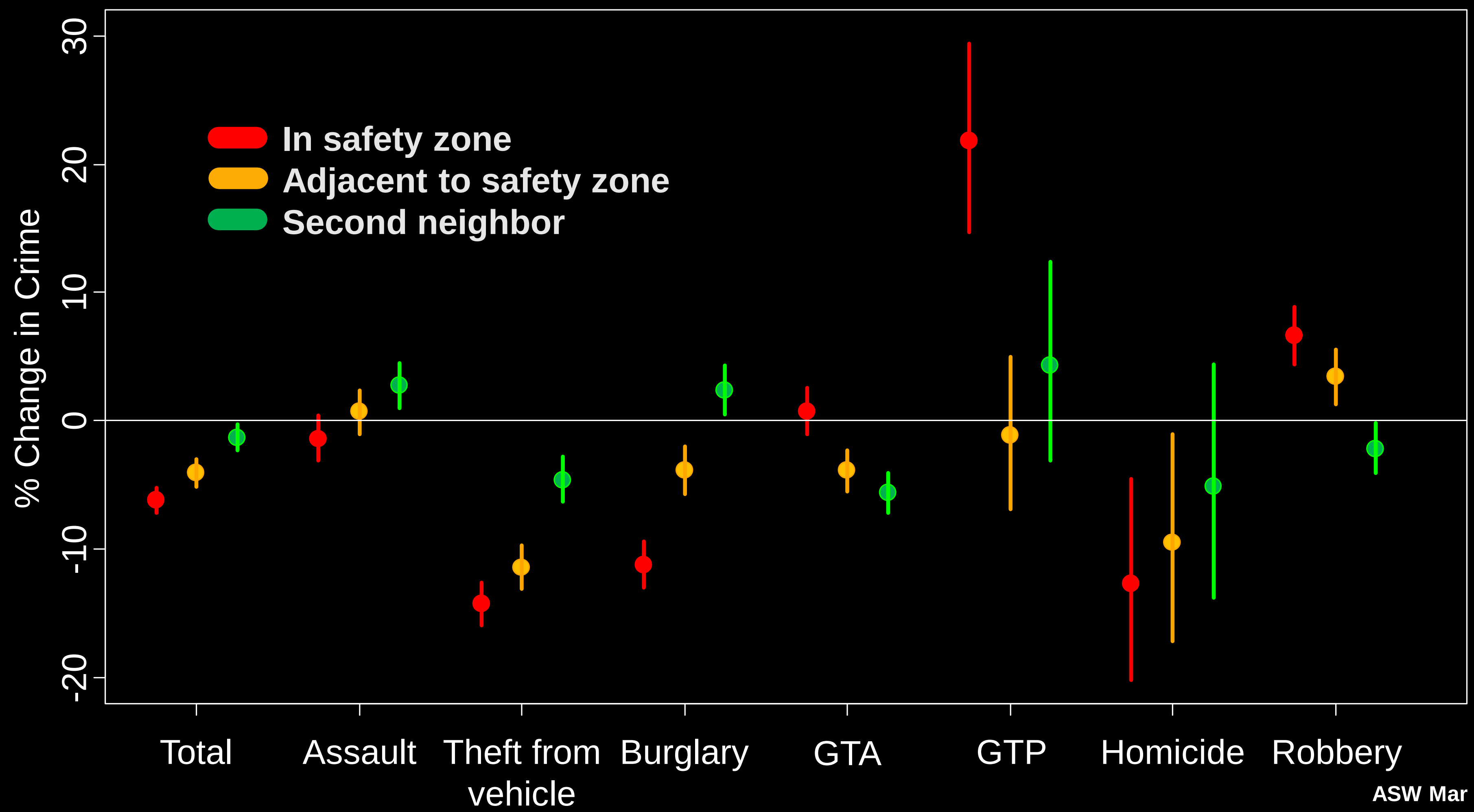
- Gang injunctions

$$\log(\lambda_{it}) = \beta_1 \mathrm{InSZ}_{it} + \beta_2 \mathrm{NextSZ}_{it} + \beta_3 \mathrm{2ndNeighbor}_{it} + \alpha_i + \gamma_t$$

- Shelters

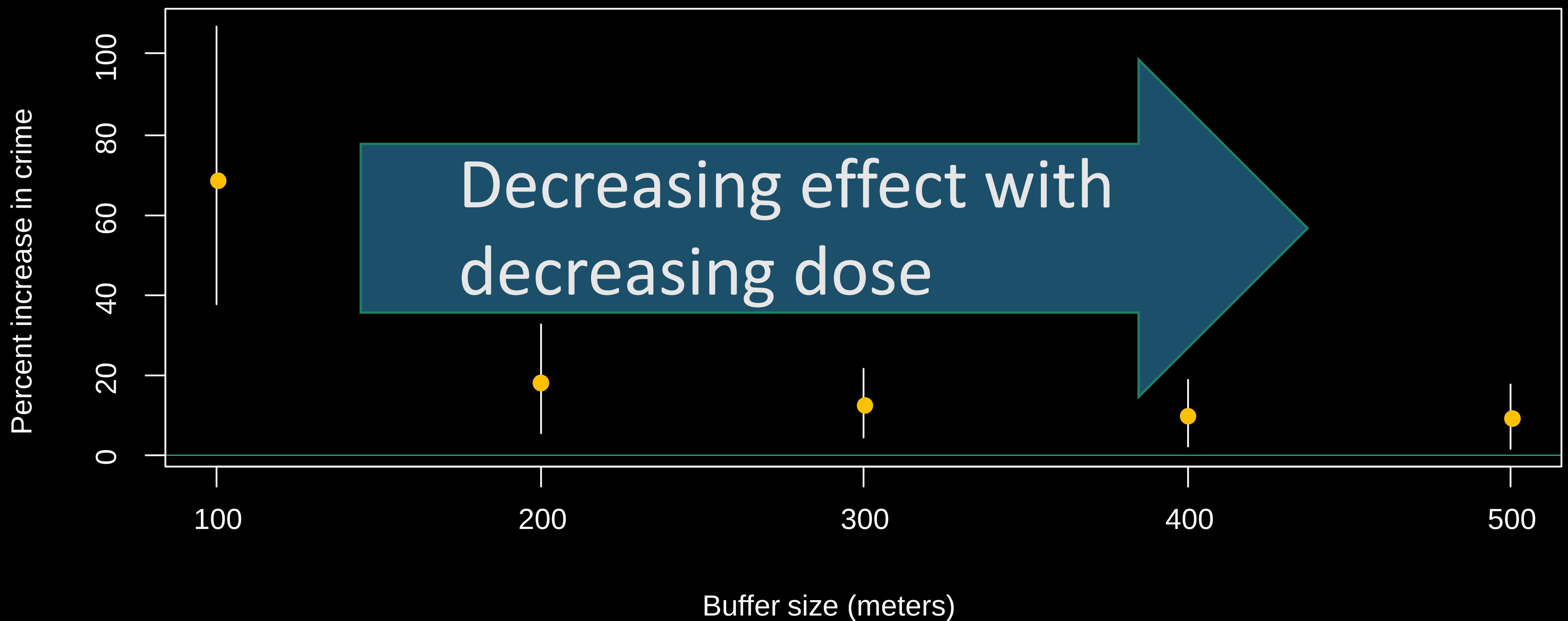
$$\log(\lambda_{it}) = \beta_1 \mathrm{shelter}_{it} + \alpha_i + \gamma_t$$

Neighborhoods Closest to Safety Zones See Largest Crime Decreases



ASW Mar 2017

Large Increase in Property Crime When Shelters Open



ASW Mar 2017

But There Are Several Open Statistical Questions

- Spatial and temporal correlation
- Overlapping treatments
- Regression to the mean
- Permutation tests
  - Wang & DeGruttola (2016) permute start times
  - Severely underpowered with some error structures
    - Size of gang injunction effect swamped
    - Size of shelter effect survives
    - Tests the strong null of no effect at all
- Solving these in a robust framework will address numerous related questions in criminology

Statistics Can Have a Prominent Role in Crime and Justice Policy

- Natural experiments
- Careful matching of groups with propensity scores and DR estimation
- Benchmarking officers
- Conditional logistic regression for case-control studies
- Stepped wedge designs for place-based interventions

Penn
University of Pennsylvania

Statistical Applications in Crime and Justice

Greg Ridgeway
Department of Criminology
Department of Statistics