Methods for Racial Profiling Analysis

Greg Ridgeway
RAND Statistics Group
Safety & Justice Research Program
Santa Monica, CA

October 26, 2008

Racial profiling is a growing concern

- I-95 “turnpike” studies in the mid-1990s raised public concern about racial profiling
- Public concern has led to widespread action
  - 26 states have passed legislation and hundreds of cities collect data
- The End of Racial Profiling Act of 2007 would mandate data collection to receive federal funds
- Should officers use racial profiling?
  - Tenth Circuit: “unequal application of criminal law to white and black persons was one of the central evils addressed by the framers of the Fourteenth Amendment”

Analytic quality is weak

- A growing number of studies claim racial profiling based on analysis of data collected
  - Texas: Concluded that “75% of agencies stop more black and Latino drivers than white drivers”
- And some studies hastily conclude no profiling occurs based on analyzed data
  - Sacramento:
    % black drivers stopped =
    % black crime suspect descriptions

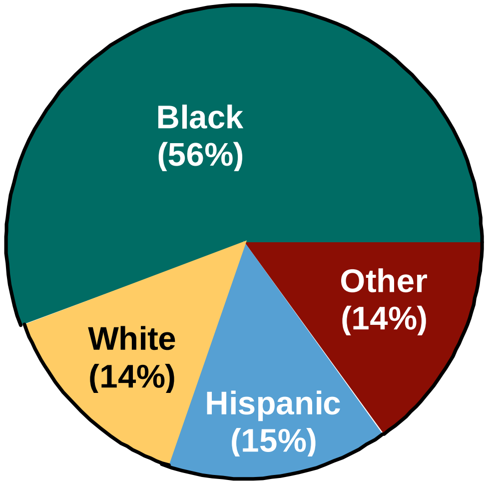
Why is testing for racial profiling so hard?

Racial Distribution of People Stopped

Racial Distribution of People at Risk of Being Stopped

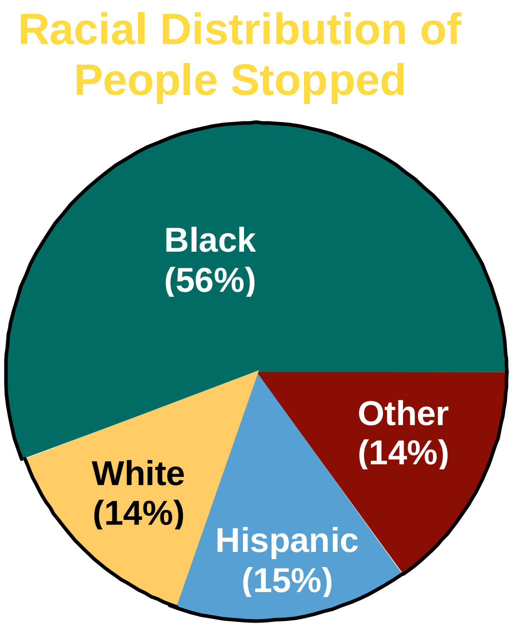
Why is testing for racial profiling so hard?

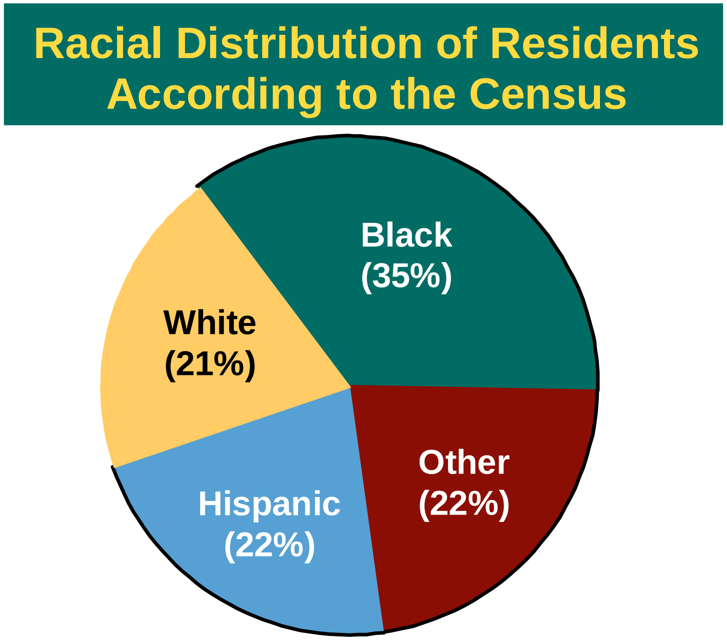
Racial Distribution of People Stopped



Racial Distribution of People at Risk of Being Stopped

Why is testing for racial profiling so hard?

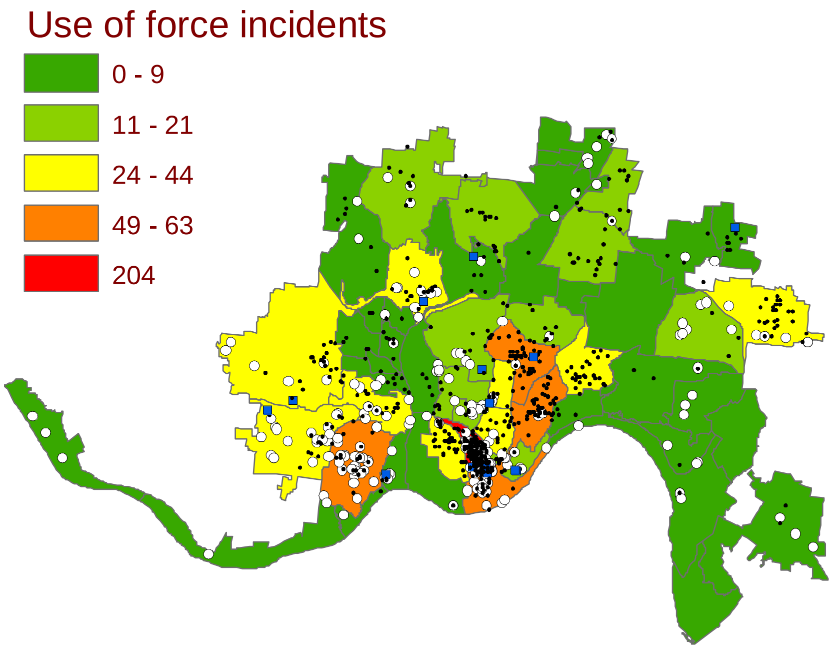




- The difference may result from:
  - A race bias
  - Car ownership, time on the road, and care
  - Exposure to police

A new approach

- Gauge department wide racial bias in the decision to stop
- Identify potential problem officers with internal benchmarking
- Assess racial bias in post-stop activity with propensity scores



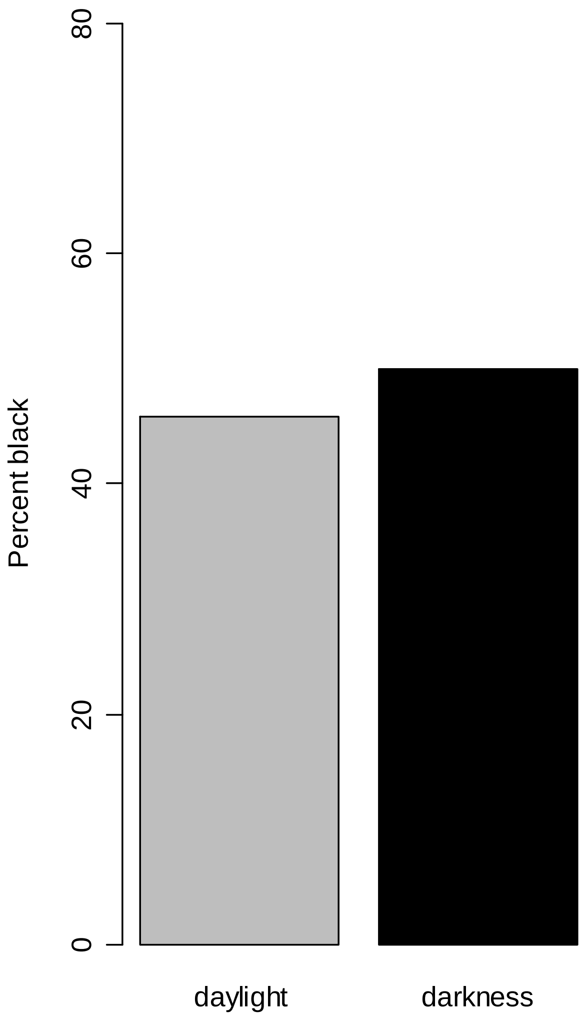
Step #1: Bias in the decision to stop

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

Central question: Does an officer's ability to identify race of driver in advance influence which drivers he stops?

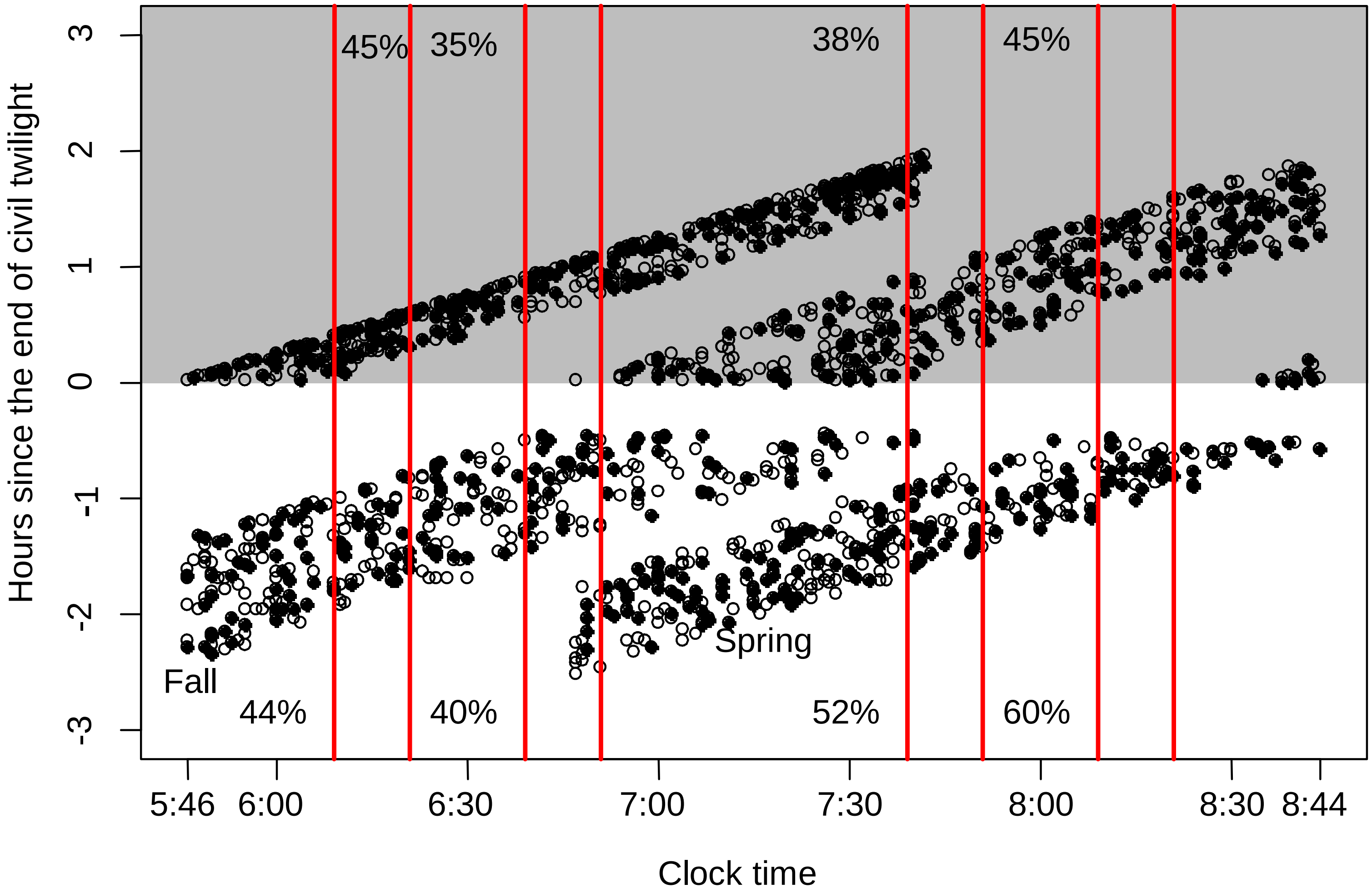
- The ability to discriminate requires officers identifying the race in advance (e.g. Goldin & Rouse, bias in orchestra auditions)
- The ability to identify race in advance of the stop decreases as it becomes dark
- We directly test whether the ability to identify the race affects the race distribution of the stopped drivers

Simple veil of darkness test



- CPD officers stop a greater proportion of black drivers at night than during the day
- This is counter to the racial profiling hypothesis

Adjusting for “clock time”



Development of the test

- In the absence of a race bias $K(t) = 1$

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

- Bayes' Theorem and some algebra yield

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

Accommodate underreporting

- There is some potential underreporting

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

$$\log K(t) =$$

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

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

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

Decomposition of the race effect

$$\log K(t) = \text{stop distribution} + \text{exposure} + \text{reporting}$$

- We can estimate the stop ratio using logistic regression

$$\log \frac{P(B|R,S,d,t)}{1-P(B|R,S,d,t)} = \beta_0 + \beta_1 d + g(t)$$

- $g(t)$ is some flexible function of t (e.g. $t + t^2 + t^3$)
- Assume exposure and reporting terms are 0
- $\log K(t) = -\beta_1$

Results: VoD estimates of bias, all months

Year	$K(t)$	95% interval	N
2003	1.04	(0.90,1.20)	3,899
2004	0.99	(0.87,1.14)	4,346
2005	1.06	(0.94,1.20)	5,193
2006	0.90	(0.79,1.02)	4,644
Combined	0.99	(0.93,1.06)	18,082

- Includes all stops during the evening intertwilight period

Results: VoD estimates of bias, Daylight Savings Time

Year	$K(t)$	95% interval	N
2003	1.02	(0.70,1.47)	543
2004	1.19	(0.80,1.77)	465
2005	1.10	(0.81,1.51)	763
2006	0.71	(0.51,1.00)	606
Combined	0.98	(0.82,1.16)	2,377

- Includes all stops occurring within four weeks of the spring or fall Daylight Saving Time change during the evening intertwilight period

Step #2: Internal benchmarking

G. Ridgeway and J.M. MacDonald. "Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops."

- 83% of this officer's stops involve a black driver

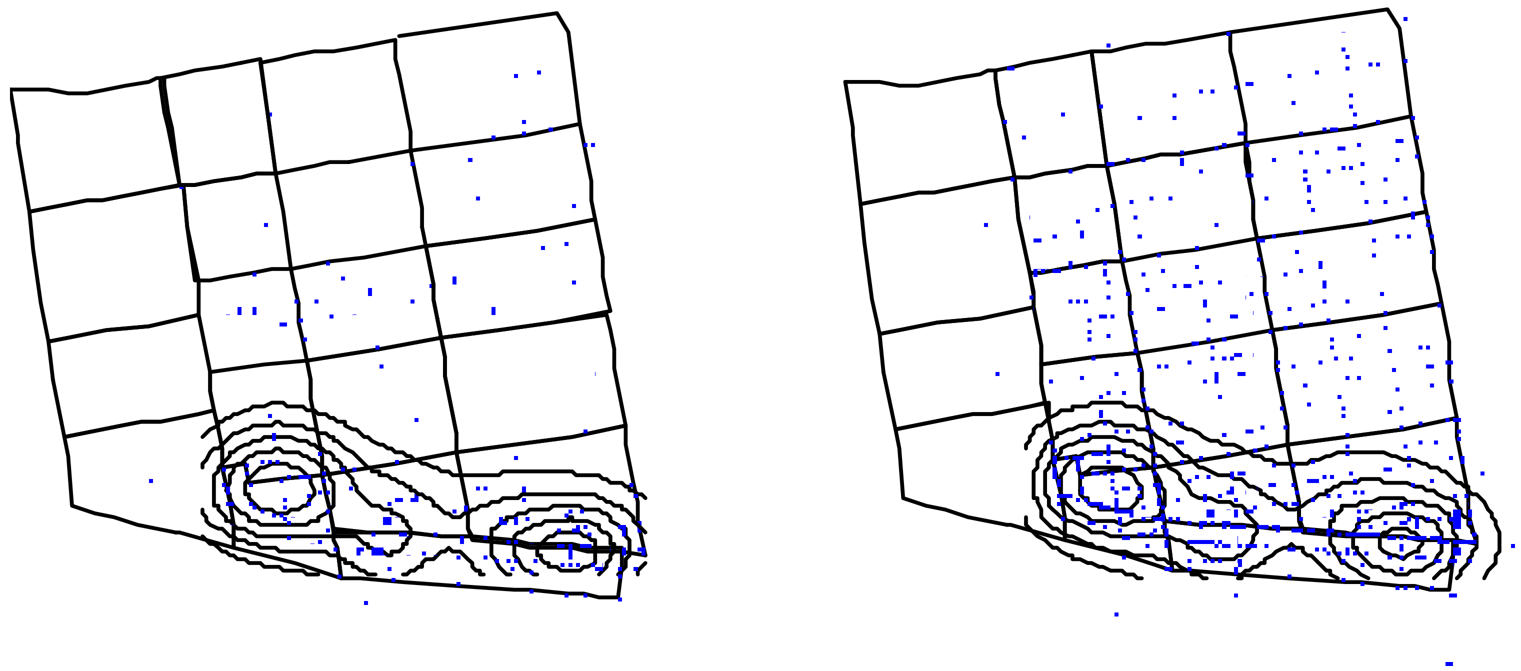
Stop Characteristic		Example Officer (%) ($n = 392$)
Month	January	3
	February	4
	March	8
Day of the week	Monday	13
	Tuesday	11
	Wednesday	14
Time of day	(4-6 p.m.]	9
	(6-8 p.m.]	8
	(8-10 p.m.]	23
	(10 p.m. -12 a.m.]	17
Patrol borough	Brooklyn North	100
Precinct	B	98
	C	1
Outside		96
In uniform	Yes	99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

Internal benchmark

- 78% of similarly situated stops made by other officers involved black drivers

Stop Characteristic		Example Officer (%) ($n = 392$)	Internal Benchmark (%) ($ESS = 3,676$)
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

Stop locations are well matched



- The internal benchmarking method also matches on the higher dimensional margins

Propensity score weighting

- Reweight stops that other officers made so that they have the same distribution of features

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

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

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

  where $p(\mathbf{x})$ is the probability that a stop with features $\mathbf{x}$ involves the officer in question

- Estimate $p(\mathbf{x})$ using a flexible, non-parametric version of logistic regression
- Compare the percentage of black drivers among the officer's stops with the weighted percentage of black drivers among other stops using weights $w_i = p(\mathbf{x}_i)/(1-p(\mathbf{x}_i))$

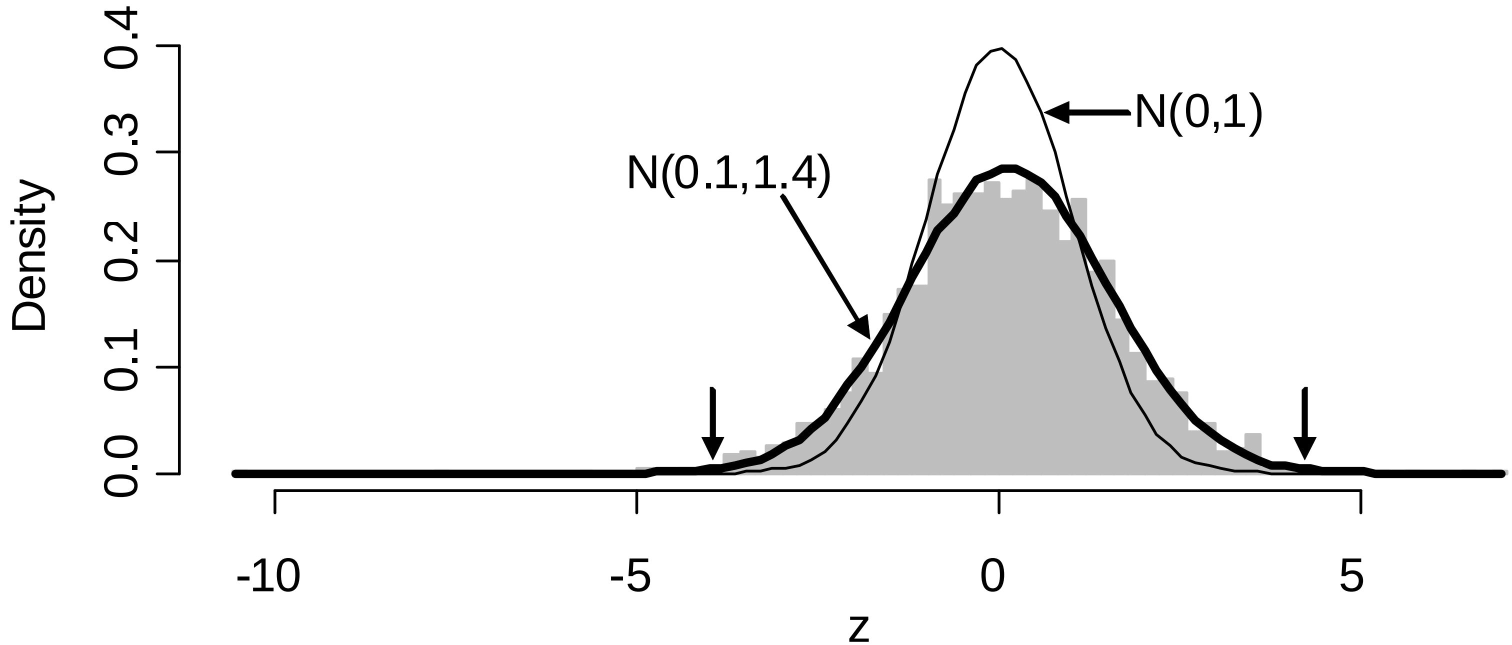
Common approach

- A common approach is to compute z-statistics for each officer

$$z = \frac{p_t - p_c}{\sqrt{\frac{p_t(1-p_t)}{n_t} + \frac{p_c(1-p_c)}{ESS}}}$$

- In the absence of racial bias this would be distributed N(0,1) and a cutoff of 2.0 would be reasonable
- With 2,756 officers and 2,756 correlated zs an appropriate reference distribution can be much wider (Efron 2006).

Estimating the false discovery rate



- Estimate $f_0(z)$ and $f(z)$ from the observed zs
- 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

Flagged officers show large disparities

Black (%)		Stops (n)		
Officer	Benchmark	Officer	Benchmark	fdr
86	55	151	773	0.03
85	67	218	473	0.38
77	56	237	1,081	0.14
75	51	178	483	0.22
64	20	59	695	0.02

Several current systems have statistical flaws

- LAPD's TEAMS II Risk Management Information System
- Pittsburgh's Performance Assessment and Review System
- Cincinnati's Risk Management System
- Phoenix's Personnel Assessment System

Step #3: Assessing race bias post-stop

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

- Central question: Are black drivers more/less likely to be cited, have long stop durations, or be searched?

Stop feature	% Black drivers (N=3,703)	% Nonblack drivers (N=3,033)
Region		
A	32%	14%
Time of day		
12am-4am	16%	8%
Resident	76%	64%
Age		
18-29	47%	38%
Reason		
Mechanical/ Registration	26%	16%
Male	75%	74%

Reweighting balances the group

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

Stop feature	% Black drivers (N=3,703)	% Nonblack drivers weighted (ESS=1,689.2)	% Nonblack drivers (N=3,033)
Region			
A	32%	33%	14%
Time of day			
12am-4am	16%	16%	8%
Resident	76%	76%	64%
Age			
18-29	47%	48%	38%
Reason			
Mechanical/ Registration	26%	26%	16%
Male	75%	76%	74%

Results: Cincinnati stop duration

Year	Stop Duration (Minutes)	Black Drivers	Nonblack (reweighted)	Nonblack (unweighted)
2003	n =	16,708	4,881	18,548
	(0,10)	40%	43%	56%
2004	n =	18,721	5,190	20,390
	(0,10)	40%	44%	59%
2005	n =	15,571	4,965	20,431
	(0,10)	45%	47%	60%
2006	n =	15,557	3,358	18,458
	(0,10)	47%	47%	56%

- Black drivers in 2006 were three times more likely to have an invalid license than white drivers (18% vs. 5%)

Results: Cincinnati search rates

Year	Discretion (Minutes)	Black Drivers	Nonblack (reweighted)	Nonblack (unweighted)
2003	n =	16,708	4,881	18,548
	High	5.9%	5.4%	2.8%
	Low	8.1%	5.5%	2.7%
2004	n =	18,721	5,190	20,390
	High	6.7%	6.2%	3.2%
	Low	10.7%	7.0%	3.9%
2005	n =	19,375	6,141	25,163
	High	6.1%	5.2%	2.8%
	Low	4.4%	3.5%	1.6%
2005	n =	20,146	5,365	24,383
	High	6.1%	6.7%	3.0%
	Low	4.9%	3.9%	1.8%

- Hit rates for black and white drivers are about 23% for high discretion searches

Summary

- Racial profiling analyses have generally confused the issue by studying irrelevant comparisons
- Credible and relevant comparisons are not difficult
  - Assess whether the ability to identify race in advance influences who gets stopped
  - Compare similarly situated officers
  - Equalize race groups on the obvious features on which they might legitimately differ

For more information...

Complete reports and papers are available at `www.rand.org`