Criminal Justice System Benchmarking

An Approach for Studying Sentencing Disparities in New York State

Greg Ridgeway, Ph.D.

Departments of Criminology and Statistics

Oct 2015

Benchmarking: The process of comparing the performance of one entity to the performance of comparison entities

Internal Benchmarking:

Benchmarking against comparison entities

Benchmarking Potential for Counties

Entities	Performance measure	Potentially confounding features
County	Sentencing	Arrestee features, case characteristics

Benchmarking Potential is Everywhere

Entities	Performance measure	Potentially confounding features
County	Sentencing	Arrestee features, case characteristics
Neighborhood	Trust in justice system	Home ownership, race, social cohesion
Police officer	Race, force	Time, place, context
Police station	Workforce size, composition	Demand for service, mix of crime types
Public defender	Caseload, time per case	Charge type, case complexity, prosecutor practice, judicial process, jail procedure
School	Truancy	Family structure, socioeconomic status

Benchmarking Police Regional Performance

Oct 2015

In Which Neighborhoods Are Police Underperforming?

- Cincinnati Police Department sponsored a citywide survey of citizens
  - Citizen satisfaction with the police
  - Perceptions of racially discriminatory police practices
  - Whether residents felt that they had personally experienced racial profiling
- 6,000 residents in Cincinnati selected via random-digit dialing and list-assisted sampling methods
  - Stratified to cover 45 defined Cincinnati neighborhoods
  - Respondents were 18 years or older

Respondents Differ on Key Features Associated with Police Satisfaction

Respondent characteristics	Respondents from Over-the-Rhine (N=146)	Respondents from neighborhoods (N=5,671)	
Less than HS	21	10	
College degree+	23	33	
Black	66	42	
White	30	53	
$20,000 or less	47	25	
$100,000 or more	6	11	
Employed (%)	60	58	
Married (%)	15	38	
Male (%)	43	36	
Age 22-29	16	8	
Age 65+	13	25	
Home owner (%)	20	60	
Children at home (%)	40	31	

Constructed Benchmark Matches Neighborhoods on These Features

Respondent characteristics	Respondents from Over-the-Rhine (N=146)	Respondents from neighborhoods (N=5,671)	Weighted respondents from other neighborhoods (N=422)
Less than HS	21	10	**21**
College degree+	23	33	**22**
Black	66	42	**65**
White	30	53	**32**
$20,000 or less	47	25	**45**
$100,000 or more	6	11	**5**
Employed (%)	60	58	**58**
Married (%)	15	38	**16**
Male (%)	43	36	**42**
Age 22-29	16	8	**17**
Age 65+	13	25	**13**
Home owner (%)	20	60	**21**
Children at home (%)	40	31	**38**

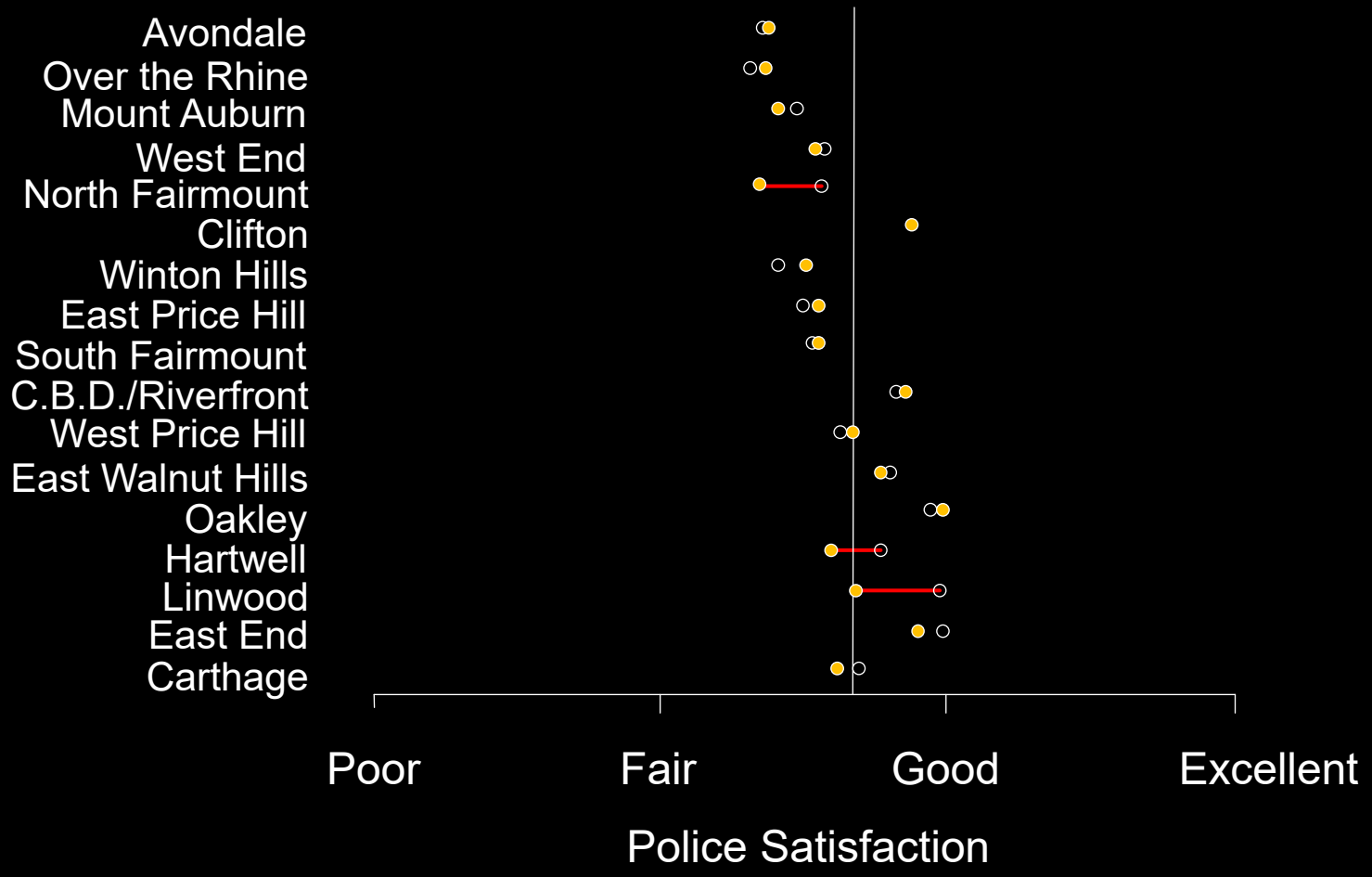
Respondents Also Match on Perceived Neighborhood Environment

Respondent characteristics	Respondents from Over-the-Rhine (N=146)	Respondents from other neighborhoods (N=5,671)	Weighted respondents from other neighborhoods (N=422)
Disorder (scale)	2.83	2.11	**2.84**
Fear of crime (scale)	2.64	2.42	**2.69**
Neighbors get together (scale)	2.38	2.41	**2.38**
Neighbors trust each other (scale)	2.13	2.78	**2.11**
Participate in neighborhood (%)	34	28	**32**
Know an officer (%)	46	32	**45**

Police Satisfaction in Over-the-Rhine Is Close to Its Benchmark

Respondent characteristics	Respondents from Over-the-Rhine (N=146)	Respondents from neighborhoods (N=422)
Satisfaction with the Police	2.37	**2.31**
Perception of Racial Profiling	2.59	**2.65**
Personal Racial Profiling Experience	32%	**30%**

Few Neighborhoods Differ from Benchmarks



Benchmarking Police Officers

Oct 2015

Is Stopping 86% Black Surprising?

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

We Know a Lot About Stop Context

Stop Characteristic		Example Officer (%) n = 392	
	Percent black	**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	

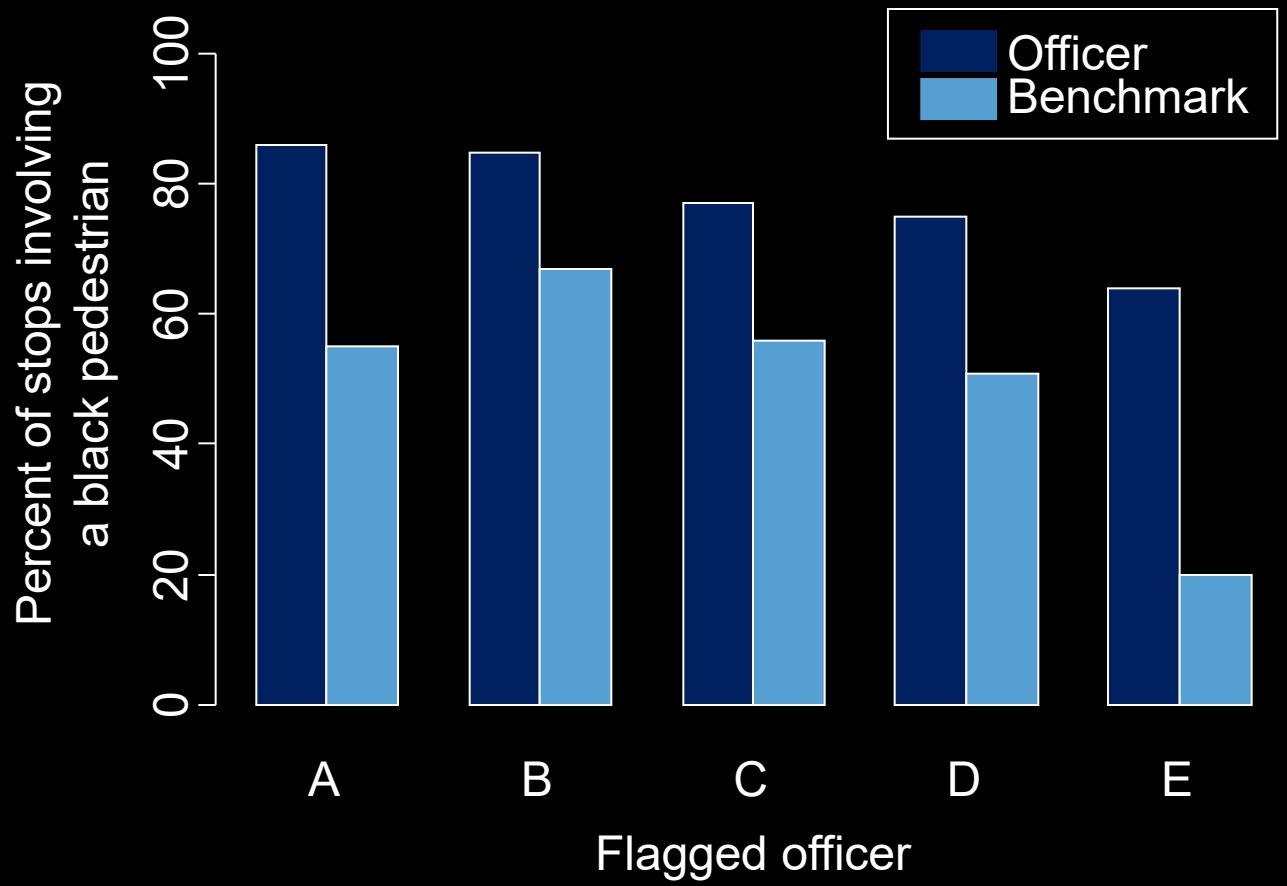
Other Officers Work in the Same Context

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

Officer's Stop Pattern Greatly Exceeds the Benchmark

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

NYPD Analysis Flagged Five Officers



Oct 2015

Benchmarking County Sentencing Disparities

New York Sentencing Study Will Use the Same Benchmarking Approach

	Cincinnati Police Satisfaction	New York City Policing	New York State Sentencing
Entity	A Cincinnati neighborhood	An NYPD officer	A New York state county
Outcome			
Benchmark			

New York Sentencing Study Will Use the Same Benchmarking Approach

	Cincinnati Police Satisfaction	New York City Policing	New York State Sentencing
Entity	A Cincinnati neighborhood	An NYPD officer	A New York state county
Outcome	Satisfaction with the police as measured by a survey	Percentage of black pedestrians stopped as recorded on UF250s	Racial difference in rate of incarceration versus non-incarceration alternatives
Benchmark			

New York Sentencing Study Will Use the Same Benchmarking Approach

	Cincinnati Police Satisfaction	New York City Policing	New York State Sentencing
Entity	A Cincinnati neighborhood	An NYPD officer	A New York state county
Outcome	Satisfaction with the police as measured by a survey	Percentage of black pedestrians stopped as recorded on UF250s	Racial difference in rate of incarceration versus non-incarceration alternatives
Benchmark	Residents in other Cincinnati neighborhoods matched on demographics, neighborhood perception	Stops made by other NYPD officers at the same times, places, contexts	Offenders charged in other New York counties with similar case features

Analysis Will Benchmark Counties on Sentencing Outcomes by Race

- Entity: Each county in New York state
- Outcome: Sentencing by race
  - Dismissal, community service, fines, probation terms, prison terms
  - Compare rates of incarceration versus non-incarceration alternatives between racial/ethnic groups
- Benchmark: Similar offenders arrested/charged in other counties
  - Match on case features: crime severity, criminal history

DCJS Will Provide Data on Arrests, Charges, Dispositions, and Sentences

- DCJS providing
  - Top Charge Criminal History file of all records reaching final disposition in 2000 through 2014
  - All underlying arrest charges associated with the arrests in the Criminal History top charge file
  - All arraignment charges associated with the arrests in the Criminal History top charge file
  - All disposition charges associated with the arrests in the Criminal History top charge file, including sentencing information for charges with a conviction disposition

Timeline

- As of September 21, 2015 I do not have DCJS data
  - DCJS has prepared a draft project description, but needs more time to prepare the data
- Draft by December 1, 2015
- Final by April 1, 2015

Criminal Justice System Benchmarking

An Approach for Studying Sentencing Disparities in New York State

Greg Ridgeway, Ph.D.

Departments of Criminology and Statistics

Oct 2015

Benchmarking Police with Synthetic Controls

Oct 2015

Benchmarking the Israel Police

- The Israel Police compare themselves to
  - The United States as a whole
  - New York City
  - Netherlands
- None of these match on obvious factors
  - Crime burden, Israel has a low homicide rate and high auto theft rate
  - Population under 18

Benchmark for Be'er Sheva Station Uses Communities Not Previously Considered

City	Weight	Population
Chula Vista, California	0.18	215,447
Gilbert, Arizona	0.17	204,904
Vancouver, Washington	0.14	160,826
Henderson, Nevada	0.13	246,369
Scottsdale, Arizona	0.09	233,105
Port St. Lucie, Florida	0.07	145,740
Boise City, Idaho	0.06	202,703
Warren, Michigan	0.05	133,941
Pueblo, Colorado	0.05	103,626
Overland Park, Kansas	0.03	170,216
Casa Grande, Arizona	0.03	38,667

Synthetic Police Station Matches on Key Features

Statistic	Be'er Sheva	U.S. Benchmark
Population (000)	193.4	185.45
percent < 18	0.26	0.26
percent males < 18	0.27	0.27
percent > 65	0.12	0.12
Homicide/1,000 population	0.01	0.01
Robbery/1,000 population	0.73	0.80
Assault/1,000 population	1.51	1.51
Burglary/1,000 population	6.03	5.66
Auto Theft/1,000 population	5.23	4.64
Theft/1,000 population	17.62	18.43

What Would American Style Policing Look Like in Be'er Sheva?

Category	U.S. Benchmark	Be'er Sheva Authorized	Be'er Sheva Actual
Sworn full-time	267	259	235
Civilian full-time	113	0	0
Patrolmen	144	115	96
Investigators	54	65	62
Intelligence		40	38
Traffic		14	14
Community police		13	13
Logistics		6	6
Command		6	6

Benchmark Combines Three Modern Statistical Techniques

- Propensity score weighting
- Double robust estimation
- False discovery rate

Propensity Score Weighting

- Propensity scores reweight the other officer's stops to resemble the target officer's stops

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

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

$$= \frac{p(\mathbf{x})}{1-p(\mathbf{x})} K$$

- Estimate $p(\mathbf{x})$ using a flexible, non-parametric version of logistic regression

Example Internal Benchmark for an NYPD Officer

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

Double Robust Estimation

- Propensity score weighted logistic regression removes remaining observed confounding

$$\ell = \sum_{i=1}^{N} w_i \left(y_i f(t_i, \mathbf{x}_i) - \log\left(1 + \exp\left(f(t_i, \mathbf{x}_i)\right)\right)\right)$$

$$f(t, x) = \alpha + \gamma t + \beta' \mathbf{x}$$

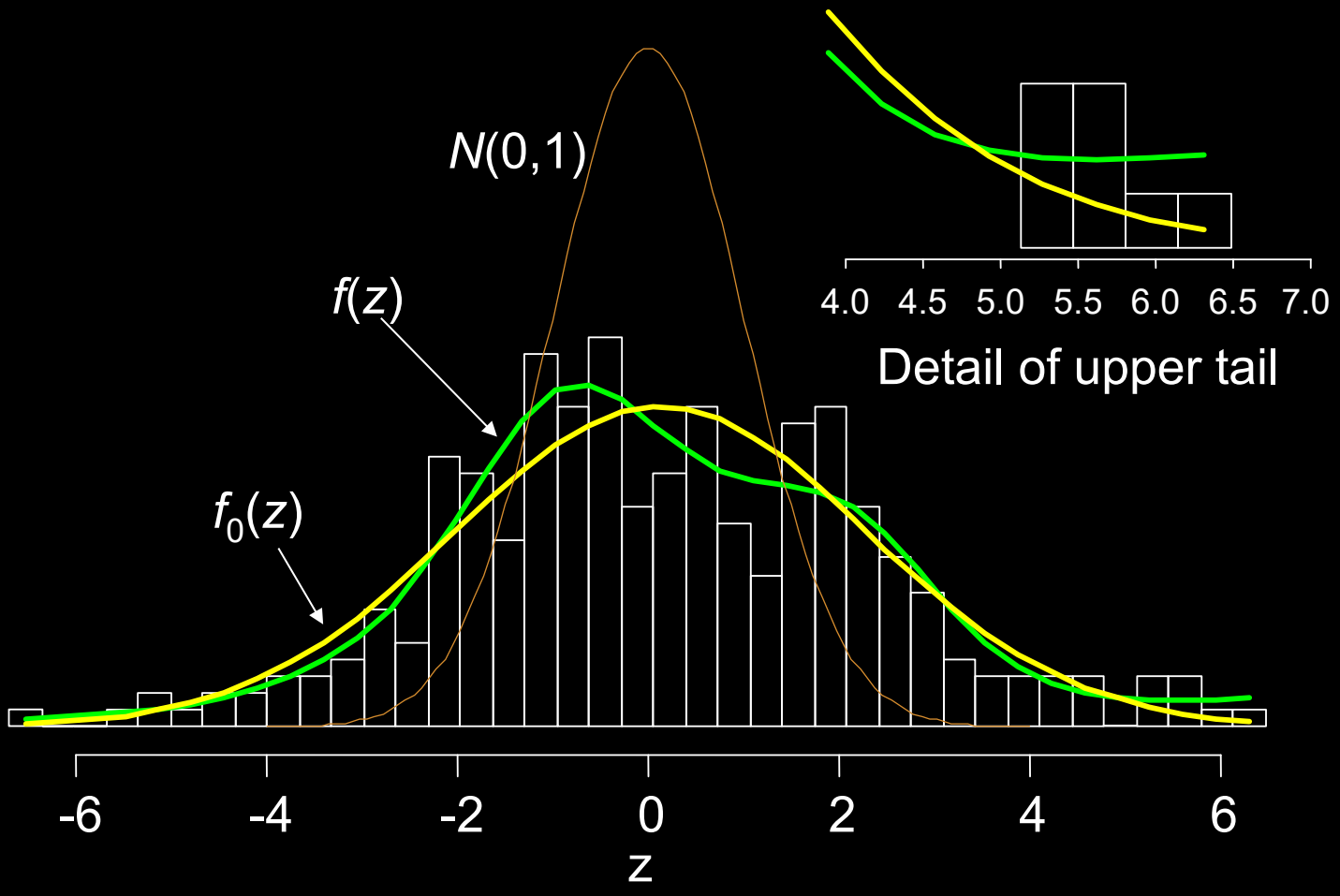
- The z-test for $\gamma = 0$ will be consistent if *either* the propensity score *or* regression model is correct

Benchmark Also Matches Fine Location Data

False Discovery Rates

$$P(\text{problem}|z) = 1 - P(\text{no problem}|z)$$
$$= 1 - \frac{f(z|\text{no problem})f(\text{no problem})}{f(z)}$$
$$\geq 1 - \frac{f_0(z)}{f(z)}$$

Use Empirical Distribution to Estimate $f_0(z)$ and $f(z)$



Cincinnati Used This Approach