Which police officers escalate force?

Greg Ridgeway

Rebecca W. Bushnell Professor of Criminology
Professor of Statistics and Data Science
University of Pennsylvania

February 18, 2026

Penn
UNIVERSITY *of* PENNSYLVANIA

Some officers seem inclined to escalate

- Laquan McDonald shooting, October 20, 2014
- CPD Officer Van Dyke fired 16 rounds
- Officer Walsh fired no rounds, holstering his firearm

Penn
UNIVERSITY of PENNSYLVANIA

Untangling officer risk from environment risk

Officers who use more force differ from other officers in obvious ways

- In the field
- In particular environments
- Conducting higher risk operations

Untangling officer risk from environment risk

Officers who use more force differ from other officers in obvious ways

- In the field
- In particular environments
- Conducting higher risk operations

“the overrepresentation of minority officers among police shooters [is] closely associated with racially varying pattern of assignment, socialization, and residence” - Fyfe (1981)

Match officers on the same scene

Find *features of officers* predictive of being a shooter (Ridgeway 2016)

- Data from 175 NYPD shooting incidents (239 shooters, 155 non-shooters)
- Conditional logistic regression
- Officers frequently accumulating negative marks in their file 3x more likely to shoot

Match officers on the same scene

Find *features of officers* predictive of being a shooter (Ridgeway 2016)

- Data from 175 NYPD shooting incidents (239 shooters, 155 non-shooters)
- Conditional logistic regression
- Officers frequently accumulating negative marks in their file 3x more likely to shoot

Find *features* predictive of rounds fired (Ridgeway, Cave, Grieco, Loeffler 2021)

- Data from 317 shootings from 46 large agencies (849 officers, 5026 rounds)
- Truncated conditional Poisson regression
- No officer feature predictive of number rounds

Match officers on the same scene

Find *features of officers* predictive of being a shooter (Ridgeway 2016)

- Data from 175 NYPD shooting incidents (239 shooters, 155 non-shooters)
- Conditional logistic regression
- Officers frequently accumulating negative marks in their file 3x more likely to shoot

Find *features* predictive of rounds fired (Ridgeway, Cave, Grieco, Loeffler 2021)

- Data from 317 shootings from 46 large agencies (849 officers, 5026 rounds)
- Truncated conditional Poisson regression
- No officer feature predictive of number rounds

Analyses did not pinpoint specific officers, results not actionable

- Identify *specific officers* rather than features of officers
- Sample size requirements demand moving beyond shootings

Penn
UNIVERSITY *of* PENNSYLVANIA

Type of force depends on officer and environment

- $Y = y$ indicates type of force, $y \in \{0, 1, 2, 3\}$
- Each officer has λ, latent propensity to escalate force
- Environment $\mathbf{z}$ (e.g., time, place, lighting, suspect, policies and laws)

$$P(Y_i = y|\mathbf{z}) \quad = \quad f(h_y(\mathbf{z}) + s_y\lambda_i)$$

Type of force depends on officer and environment

- $Y = y$ indicates type of force, $y \in \{0, 1, 2, 3\}$
- Each officer has λ, latent propensity to escalate force
- Environment $\mathbf{z}$ (e.g., time, place, lighting, suspect, policies and laws)

$$P(Y_i = y|\mathbf{z}) \quad = \quad f(h_y(\mathbf{z}) + s_y\lambda_i)$$

- Challenges
  - $\mathbf{z}$ complex and hard to measure
  - $h_y(\mathbf{z})$ also highly complex
  - No data collected when $y = 0$
  - No obvious choice for f

Three conditions determine unique f free from $h_y(\mathbf{z})$

1. $P(Y_i = y|\mathbf{z}) = f(h_y(\mathbf{z}) + s_y\lambda_i)$

Three conditions determine unique f free from $h_y(\mathbf{z})$

1. $P(Y_i = y|\mathbf{z}) = f(h_y(\mathbf{z}) + s_y\lambda_i)$
2. Increases in λ correspond to increases in relative risk of more serious force. If $y_2 > y_1$ and $\epsilon > 0$ then

$$\frac{P(Y = y_2|\mathbf{z}, \lambda + \epsilon)}{P(Y = y_1|\mathbf{z}, \lambda + \epsilon)} > \frac{P(Y = y_2|\mathbf{z}, \lambda)}{P(Y = y_1|\mathbf{z}, \lambda)}$$

Three conditions determine unique f free from $h_y(\mathbf{z})$

1. $P(Y_i = y|\mathbf{z}) = f(h_y(\mathbf{z}) + s_y\lambda_i)$
2. Increases in λ correspond to increases in relative risk of more serious force. If $y_2 > y_1$ and $\epsilon > 0$ then

$$\frac{P(Y = y_2|\mathbf{z}, \lambda + \epsilon)}{P(Y = y_1|\mathbf{z}, \lambda + \epsilon)} > \frac{P(Y = y_2|\mathbf{z}, \lambda)}{P(Y = y_1|\mathbf{z}, \lambda)}$$

3. Use conditional likelihood for officers matched on the same scene to eliminate dependence on $h_y(\mathbf{z})$

Penn
UNIVERSITY of PENNSYLVANIA

Five steps to arrive at the unique f

1. Only candidate sufficient statistic that could eliminate $h_y(\mathbf{z})$ in a conditional likelihood is $\mathbf{k}$, the counts of each force type used

Five steps to arrive at the unique f

1. Only candidate sufficient statistic that could eliminate $h_y(\mathbf{z})$ in a conditional likelihood is $\mathbf{k}$, the counts of each force type used
2. Conditional likelihood for one incident with m officers is

$$CL(\mathbf{h}, \mathbf{s}, \lambda \mid \mathbf{k}) = \frac{\prod_{i=1}^{m} f(y_i, h_{y_i}(\mathbf{z}) + s_{y_i}\lambda_i)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \prod_{i=1}^{m} f(y_i^*, h_{y_i^*}(\mathbf{z}) + s_{y_i^*}\lambda_i)}$$

Five steps to arrive at the unique f

1. Only candidate sufficient statistic that could eliminate $h_y(\mathbf{z})$ in a conditional likelihood is $\mathbf{k}$, the counts of each force type used
2. Conditional likelihood for one incident with m officers is

$$CL(\mathbf{h}, \mathbf{s}, \lambda \mid \mathbf{k}) = \frac{\prod_{i=1}^{m} f(y_i, h_{y_i}(\mathbf{z}) + s_{y_i}\lambda_i)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \prod_{i=1}^{m} f(y_i^*, h_{y_i^*}(\mathbf{z}) + s_{y_i^*}\lambda_i)}$$

3. $h_y(\mathbf{z})$ cancels out *if and only if* $f(y, h + \lambda) \; = \; g(y, h)\, G(y, \lambda)$

Five steps to arrive at the unique f

1. Only candidate sufficient statistic that could eliminate $h_y(\mathbf{z})$ in a conditional likelihood is $\mathbf{k}$, the counts of each force type used
2. Conditional likelihood for one incident with m officers is

$$CL(\mathbf{h}, \mathbf{s}, \lambda \mid \mathbf{k}) = \frac{\prod_{i=1}^{m} f(y_i, h_{y_i}(\mathbf{z}) + s_{y_i}\lambda_i)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \prod_{i=1}^{m} f(y_i^*, h_{y_i^*}(\mathbf{z}) + s_{y_i^*}\lambda_i)}$$

3. $h_y(\mathbf{z})$ cancels out *if and only if* $f(y, h + \lambda) \ = \ g(y, h)\, G(y, \lambda)$
4. Only (non-trivial) solution is f, g, and G are exponential

Penn
University *of* Pennsylvania

Five steps to arrive at the unique f

1. Only candidate sufficient statistic that could eliminate $h_y(\mathbf{z})$ in a conditional likelihood is $\mathbf{k}$, the counts of each force type used
2. Conditional likelihood for one incident with m officers is

$$CL(\mathbf{h}, \mathbf{s}, \lambda \mid \mathbf{k}) = \frac{\prod_{i=1}^{m} f(y_i, h_{y_i}(\mathbf{z}) + s_{y_i}\lambda_i)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \prod_{i=1}^{m} f(y_i^*, h_{y_i^*}(\mathbf{z}) + s_{y_i^*}\lambda_i)}$$

3. $h_y(\mathbf{z})$ cancels out *if and only if* $f(y, h+\lambda) = g(y, h)\, G(y, \lambda)$
4. Only (non-trivial) solution is f, g, and G are exponential
5. Only exponential model satisfying relative risk constraint is Anderson's ordinal stereotype model

Penn
University of Pennsylvania

Anderson's ordinal stereotype model

$$P(Y_i = y|\mathbf{z}) = \frac{\exp(\alpha_y + s_y\beta'\mathbf{x})}{\sum_{j=0}^{J}\exp(\alpha_j + s_j\beta'\mathbf{x})} \quad \text{(standard OSM)}$$

$$= \frac{\exp(h_y(\mathbf{z}) + s_y\lambda_i)}{\sum_{j=0}^{J}\exp(h_j(\mathbf{z}) + s_j\lambda_i)}$$

where $s_0 = 0, s_1 = 1 \leq s_2 \leq ... \leq s_J$

λ measures increase in relative risk

$$\frac{P(Y=y_2|\lambda=\lambda_1)}{P(Y=y_1|\lambda=\lambda_1)} = \frac{P(Y=y_2|\lambda=\lambda_0)}{P(Y=y_1|\lambda=\lambda_0)} e^{(s_{y_2}-s_{y_1})(\lambda_1-\lambda_0)}$$

- e^{λ} is the multiplicative change in the risk of Level 1 force relative to Level 0 associated with escalation λ, relative to a reference officer with $\lambda = 0$

$\frac{1}{s_{y_2}-s_{y_1}}$ measures "distance" between force levels

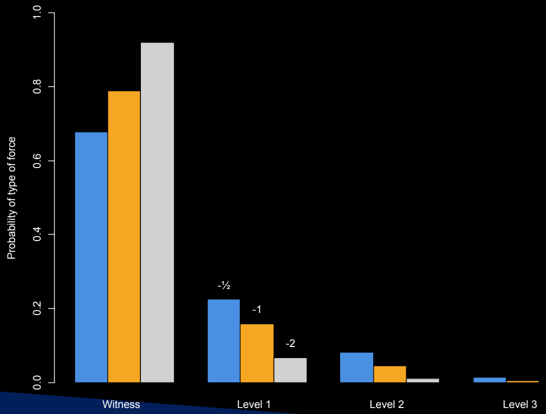
In order for the relative risk of using force type y_2 instead of force type y_1 to increase by a factor of C, λ would need to increase by

$$\frac{1}{s_{y_2} - s_{y_1}} \log C$$

Penn
UNIVERSITY *of* PENNSYLVANIA

Ordinal stereotype imposes a distribution on force types

- $h_y(\mathbf{z})$ set to match Seattle's force rate for force type y
- $\mathbf{s} = \{0, 1, \frac{3}{2}, 2\}$
- $\lambda_1 = -\frac{1}{2}, \lambda_2 = -1, \lambda_3 = -2$



Conditional likelihood for an example incident

Consider a use-of-force incident

- three officers on scene ($m = 3$)
- one officer does nothing (Level 0)
- one officer physically restrains (Level 1)
- one officer strikes baton to the head (Level 3)

What's the probability $Y_1 = 0$, $Y_2 = 1$, and $Y_3 = 3$?

Conditional likelihood for an example incident

Consider a use-of-force incident

- three officers on scene ($m = 3$)
- one officer does nothing (Level 0)
- one officer physically restrains (Level 1)
- one officer strikes baton to the head (Level 3)

What's the probability $Y_1 = 0$, $Y_2 = 1$, and $Y_3 = 3$?

$$P(Y_1 = 0, Y_2 = 1, Y_3 = 3 \,|\, \mathbf{k} = \{1, 1, 0, 1\}, \mathbf{s}, \lambda, h(\mathbf{z}))$$
$$= \frac{e^{s_0\lambda_1 + s_1\lambda_2 + s_3\lambda_3}}{e^{s_0\lambda_1 + s_1\lambda_2 + s_3\lambda_3} + \ldots + e^{s_3\lambda_1 + s_1\lambda_2 + s_0\lambda_3}}$$

Penn
University of Pennsylvania

Conditional likelihood for an example incident

Consider a use-of-force incident

- three officers on scene ($m = 3$)
- one officer does nothing (Level 0)
- one officer physically restrains (Level 1)
- one officer strikes baton to the head (Level 3)

What's the probability $Y_1 = 0$, $Y_2 = 1$, and $Y_3 = 3$?

$$P(Y_1 = 0, Y_2 = 1, Y_3 = 3 \,|\, \mathbf{k} = \{1,1,0,1\}, \mathbf{s}, \lambda, h(\mathbf{z}))$$
$$= \frac{e^{s_0\lambda_1 + s_1\lambda_2 + s_3\lambda_3}}{e^{s_0\lambda_1 + s_1\lambda_2 + s_3\lambda_3} + \ldots + e^{s_3\lambda_1 + s_1\lambda_2 + s_0\lambda_3}}$$

Probability does not depend on the environment. No need to measure $\mathbf{z}$ or worry about $h(\mathbf{z})$

Penn
UNIVERSITY *of* PENNSYLVANIA

Solves all issues... with some wrinkles

For a general incident with m officers

$$P(\mathbf{Y}=\mathbf{y}|\mathbf{s},\lambda,\mathbf{k})=\frac{\exp\left(\sum_{i=1}^{m}s_{y_i}\lambda_i\right)}{\sum_{\mathbf{y}^*\in\mathcal{K}}\exp\left(\sum_{i=1}^{m}s_{y_i^*}\lambda_i\right)}$$

- No need for environmental/situational measures
- Moments when all officers have $y=0$ have no information
- Moments involving a single officer have no information

Solves all issues… with some wrinkles

For a general incident with m officers

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{s}, \lambda, \mathbf{k}) = \frac{\exp\left(\sum_{i=1}^{m} s_{y_i} \lambda_i\right)}{\sum_{\mathbf{y}^* \in \mathcal{K}} \exp\left(\sum_{i=1}^{m} s_{y_i^*} \lambda_i\right)}$$

- No need for environmental/situational measures
- Moments when all officers have $y = 0$ have no information
- Moments involving a single officer have no information

However…

- Assumes conditional independence $\left(Y_i, Y_j \perp\!\!\!\perp \mathbf{z}, \lambda\right)$
- Denominator is computationally challenging
- λ generally not identifiable

Computational challenges

Complete conditional likelihood is

$$CL(\mathbf{s}, \lambda) = \prod_{\ell=1}^{n} \frac{\exp\left(\sum_{i=1}^{m_\ell} s_{y_{i\ell}} \lambda_{\mathrm{id}(i,\ell)}\right)}{\sum_{\mathbf{y}^* \in \mathcal{K}_\ell} \exp\left(\sum_{i=1}^{m_\ell} s_{y_i^*} \lambda_{\mathrm{id}(i,\ell)}\right)}$$

Efficient computation of Poisson-Multinomial denominator

- $m_\ell = 2$, easily computed
- $3 \leq m_\ell \leq 7$, no-repeat Heap (1967) recursive algorithm, $O(m!)$ time
- $m_\ell \geq 8$, discrete Fourier transform Lin et al (2023), $O(m^4)$ time
- Efficient denominator calculation plus parallelization by incident makes MCMC possible

λ is identifiable up to an additive constant

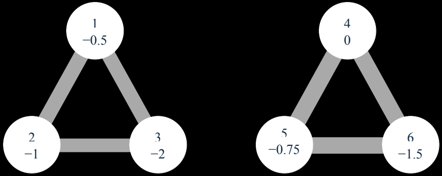
$$L_\ell(\mathbf{s}, \lambda + \mathbf{1}c) = \frac{\exp\left(\sum_{i=1}^{m_\ell} s_{y_i}(\lambda_i + c)\right)}{\sum_{\mathbf{y}^* \in \mathcal{K}_\ell} \exp\left(\sum_{i=1}^{m_\ell} s_{y_i^*}(\lambda_i + c)\right)} = L_\ell(\mathbf{s}, \lambda)$$

- Differences between two officers' λ can be identifiable

λ is identifiable up to an additive constant

$$L_\ell(\mathbf{s}, \lambda + \mathbf{1}c) = \frac{\exp\left(\sum_{i=1}^{m_\ell} s_{y_i}(\lambda_i + c)\right)}{\sum_{\mathbf{y}^* \in \mathcal{K}_\ell} \exp\left(\sum_{i=1}^{m_\ell} s_{y_i^*}(\lambda_i + c)\right)} = L_\ell(\mathbf{s}, \lambda)$$

- Differences between two officers' λ can be identifiable
- Disconnected use-of-force networks introduce additional identifiability problems

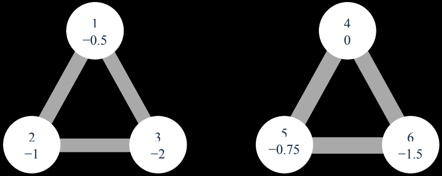


- Only officers within the same network are comparable

λ is identifiable up to an additive constant

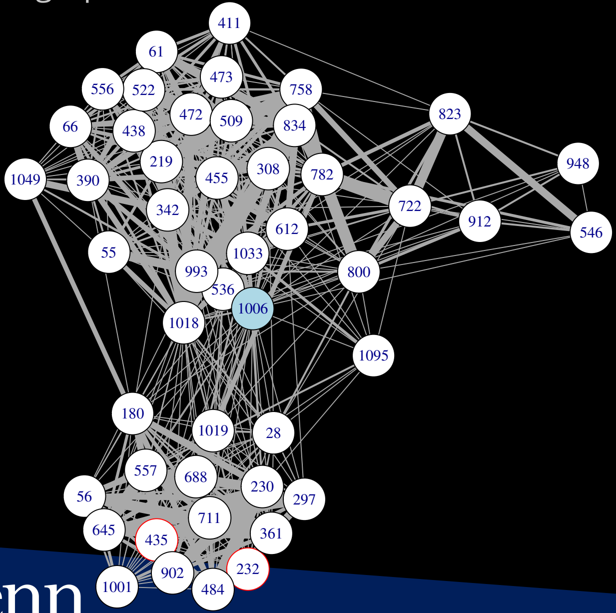
$$L_\ell(\mathbf{s}, \lambda + \mathbf{1}c) = \frac{\exp\left(\sum_{i=1}^{m_\ell} s_{y_i}(\lambda_i + c)\right)}{\sum_{\mathbf{y}^* \in \mathcal{K}_\ell} \exp\left(\sum_{i=1}^{m_\ell} s_{y_i^*}(\lambda_i + c)\right)} = L_\ell(\mathbf{s}, \lambda)$$

- Differences between two officers' λ can be identifiable
- Disconnected use-of-force networks introduce additional identifiability problems



- Only officers within the same network are comparable
- Additional complexity when comparing with officers weakly connected

Radius 1 subgraph of Officer 1006's use-of-force network



Defining peers based on conditional variance

- Use Metropolis-Hastings to draw from posterior distribution of λ
  - no identifying constraints
  - prior on λ to maintain numerical stability

Defining peers based on conditional variance

- Use Metropolis-Hastings to draw from posterior distribution of λ
  - no identifying constraints
  - prior on λ to maintain numerical stability
- Define Officer i to be in Officer 1006's local network if

$$\begin{aligned} \sigma^2_{i|1006} &= \mathrm{Var}(\lambda_i|\lambda_{1006}) \\ &\approx \mathrm{Var}(\lambda_i)(1-\rho^2_{i|1006}) < t \end{aligned}$$

- t small, comparisons involve small groups of officers, losing ability to assess department outliers
- t large, detect outliers among a large set of officers, but less precision

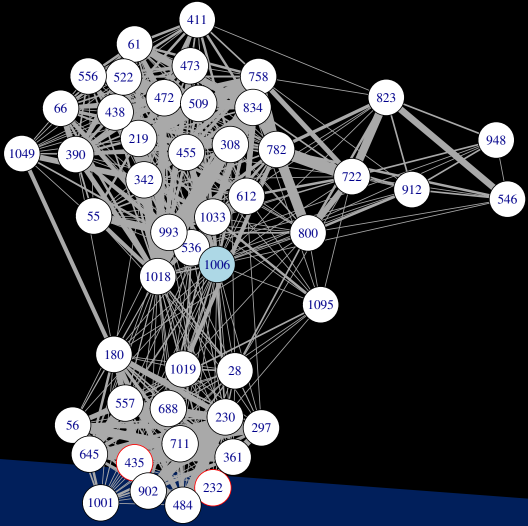
Defining peers based on conditional variance

Officer i's peer group defined as

$$\mathcal{O}_i = \{i^* \mid \mathbf{Var}(\lambda_{i^*} \mid \lambda_i, \mathbf{data}) < t\}$$

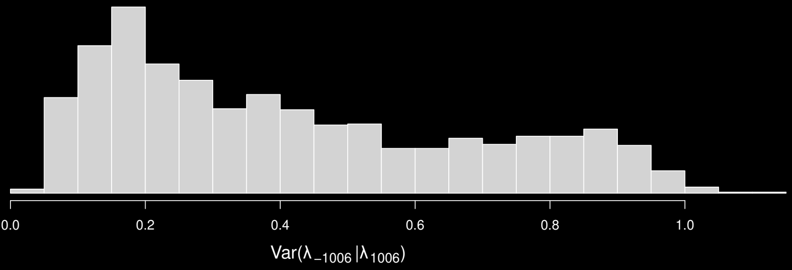
$\sigma^2_{i|1006} < 0.3$ for Officers 536, 993, 1018, 1033

$\sigma^2_{i|1006} > 0.6$ for Officers 232 and 435



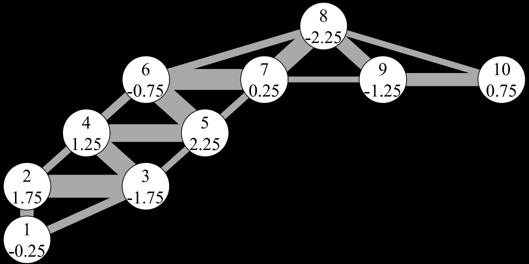
Penn
University of Pennsylvania

$\mathrm{Var}(\lambda_i \mid \lambda_{1006})$ reveals well-connected officers



- $\lambda_{1006} - \lambda_i$ is estimable with precision only if Officer 1006's and Officer i's networks are well-connected
- Officers in disconnected subgraphs or with few shared incidents will have a large conditional variance

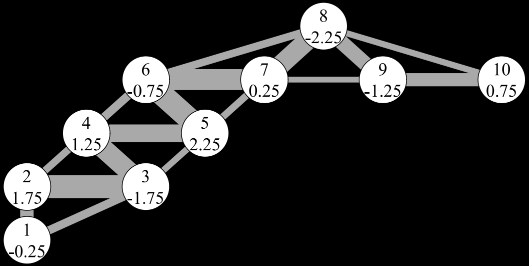
A ten-officer example (24 incidents/officer)



		posterior	
Parameter	true	mean	SD
$\lambda_1 - \lambda_2$	-2.00	-1.73	0.79
$\lambda_1 - \lambda_3$	1.50	1.74	0.81

Penn
University *of* Pennsylvania

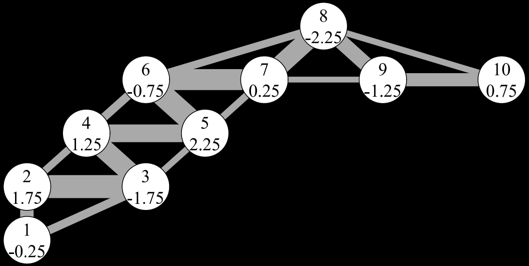
A ten-officer example (24 incidents/officer)



		posterior	
Parameter	true	mean	SD
$\lambda_1 - \lambda_2$	-2.00	-1.73	0.79
$\lambda_1 - \lambda_3$	1.50	1.74	0.81
$\lambda_1 - \lambda_4$	-1.50	-0.86	0.79
$\lambda_1 - \lambda_5$	-2.50	-1.29	0.90
$\lambda_1 - \lambda_6$	0.50	0.76	0.82
$\lambda_1 - \lambda_7$	-0.50	-0.40	0.90
$\lambda_1 - \lambda_8$	2.00	1.53	0.86
$\lambda_1 - \lambda_9$	1.00	0.91	0.88
$\lambda_1 - \lambda_{10}$	-1.00	0.16	0.92
s_2	1.50	2.25	0.67
s_3	2.00	2.56	0.68

Penn
UNIVERSITY of PENNSYLVANIA

A ten-officer example (250 incidents/officer)



		posterior	
Parameter	true	mean	SD
$\lambda_1 - \lambda_2$	-2.00	-1.53	0.26
$\lambda_1 - \lambda_3$	1.50	1.23	0.24
$\lambda_1 - \lambda_4$	-1.50	-1.48	0.36
$\lambda_1 - \lambda_5$	-2.50	-2.14	0.44
$\lambda_1 - \lambda_6$	0.50	0.59	0.41
$\lambda_1 - \lambda_7$	-0.50	-0.63	0.43
$\lambda_1 - \lambda_8$	2.00	2.16	0.47
$\lambda_1 - \lambda_9$	1.00	1.16	0.48
$\lambda_1 - \lambda_{10}$	-1.00	-0.66	0.55
s_2	1.50	1.53	0.11
s_3	2.00	2.08	0.16

Analysis of Seattle PD data

Data include

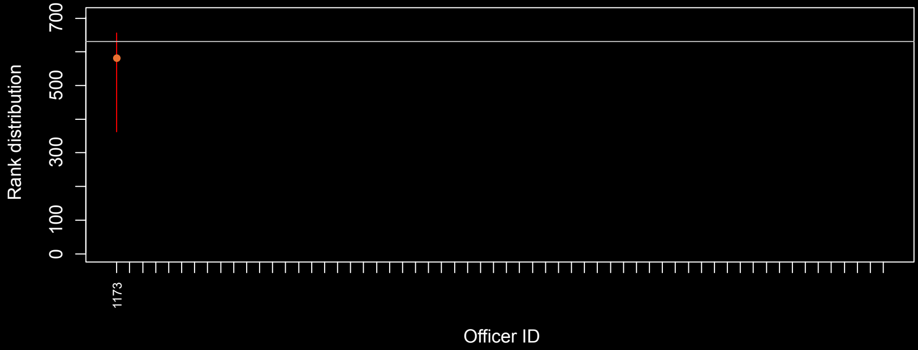
- 4,821 force incidents
- 1,503 unique officers
- 8,209 uses of force
- 28,807 officers witnessing force

Did not include

- 808 incidents with a single officer
- 455 incidents in which all used same force level
- Incidents with more than 19 officers (<8% of incidents, some had 100+)

Officer 1006 likely a high force escalator

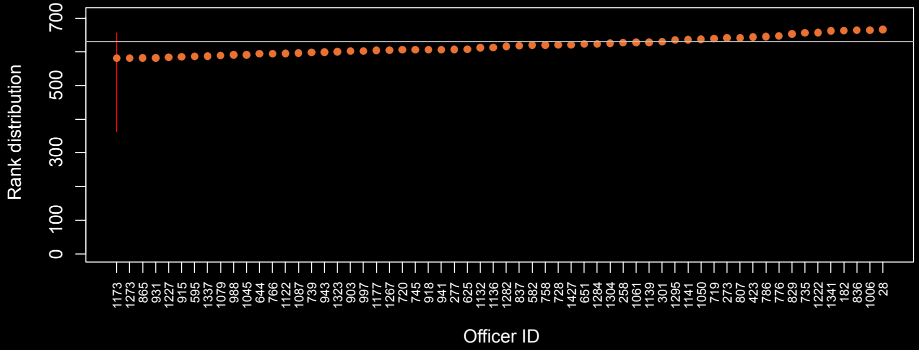
- Posterior rank distribution of λs for Officer 1006's local network
- 663 officers in Officer 1006's local network



Penn
University of Pennsylvania

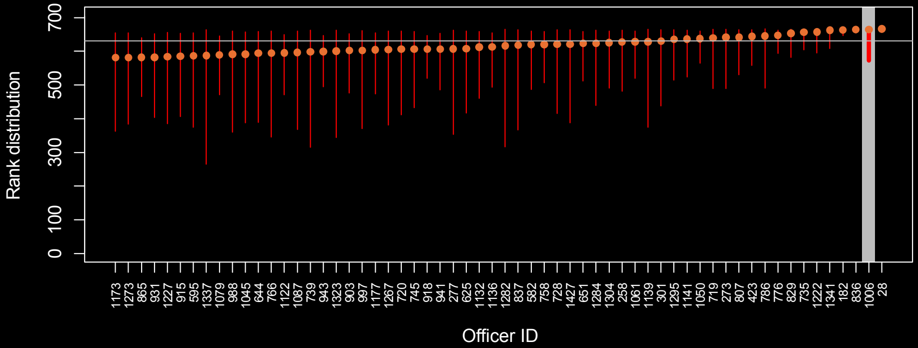
Officer 1006 likely a high force escalator

- Posterior rank distribution of λs for Officer 1006's local network
- 663 officers in Officer 1006's local network



Officer 1006 likely a high force escalator

- Posterior rank distribution of λs for Officer 1006's local network
- 663 officers in Officer 1006's local network



Flag officers with high force escalation

ID	Peers	Incidents	Count by force level				Prob. rank
			Witness	1	2	3	top 5%
#1006	663	9	1	8	0	0	0.92

- Peers is the number of officers with $\mathrm{Var}(\lambda_i \mid \lambda_{1006}, \mathbf{data}) < 0.3$
- Incidents is number of incidents involving Officer 1006
- Posterior probability is 0.92 that Officer 1006 is in the highest 5% of 663 peers
- Officer 1006 has not used serious forms of force, but is unusually hands on

Penn
UNIVERSITY of PENNSYLVANIA

Flag officers with high force escalation

ID	Peers	Incidents	Count of force level				Prob. rank	$\hat{\lambda}$	95% interval
			Witness	1	2	3	top 5%		
#1006	663	9	1	8	0	0	0.92	2.02	(0.83, 3.23)

- $\hat{\lambda}$ is centered on the peer group average λ for identifiability
- $\hat{\lambda} = 2$ implies that compared to the average peer
  - RR of Level 1 instead of Level 0 is 7x greater
  - RR of Level 2 instead of Level 1 is 11% greater
  - RR of Level 3 instead of Level 2 is 17% greater

Penn
University of Pennsylvania

Flag officers with high force escalation

ID	Peers	Incidents	Count of force level				Prob. rank	$\hat{\lambda}$	95% interval
			Witness	1	2	3	top 5%		
#836	661	75	33	39	2	1	1.00	1.99	(1.51, 2.48)
#28	661	37	12	24	1	0	1.00	2.24	(1.57, 2.93)
#182	661	79	25	43	11	0	1.00	1.88	(1.38, 2.40)
#1341	661	29	8	12	9	0	0.94	1.82	(1.01, 2.61)
#1006	663	9	1	8	0	0	0.92	2.02	(0.83, 3.23)
#735	661	49	22	17	9	1	0.90	1.58	(0.99, 2.17)
#1222	661	40	16	21	3	0	0.88	1.59	(0.93, 2.22)
#1450	662	7	0	6	1	0	0.87	1.99	(0.66, 3.34)
#829	661	52	24	15	12	1	0.82	1.49	(0.87, 2.12)

Conclusions

- Solves a long-standing problem of confounding by assignment
- Can be integrated into police early intervention systems
- Demonstrates the value of documenting witness officers, now mandated in some consent decrees such as in Chicago
- My favorite kind of problem: integration of policing, statistics, mathematics, and computer science

Which police officers escalate force?

Greg Ridgeway

Rebecca W. Bushnell Professor of Criminology
Professor of Statistics and Data Science
University of Pennsylvania

February 18, 2026

Penn
UNIVERSITY of PENNSYLVANIA