Integrating U.S. Gun Violence Data

Elizabeth Chase
Feb. 11, 2021

Problem

1997 Dickey Amendment

"None of the funds made available
for injury prevention and control
at the CDC may be used to
advocate or promote gun control."

FBI Supplemental Homicide Reports

- Uses voluntarily-reported information from police departments to identify homicides
- Mixed coverage of states 1980-2017
- Provides grouped counts

CDC Injury Data

- Uses information from National Vital Statistics System (mandatory-reported from hospitals) to identify both fatal and non-fatal injuries
- Covers all 50 states 2001-2017
- Provides grouped counts; cells with 1-9 incidents are suppressed

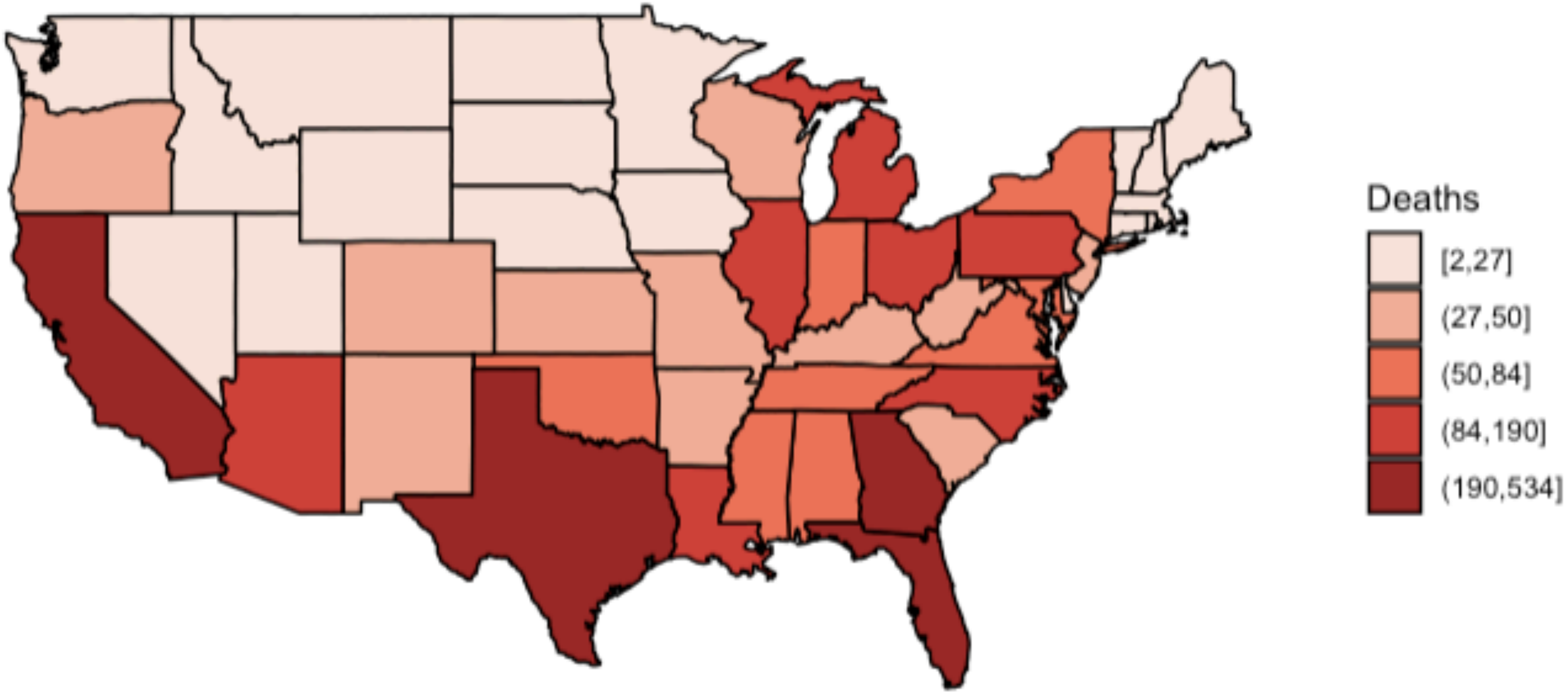
NVDRS

- Combines voluntarily-reported police records, mandatory-reported hospital records, and death records
- Has data on 36 states + DC from 2003-2017 (only 7 states with data going back to 2003)
- Provides grouped counts; cells with 1-9 incidents are suppressed

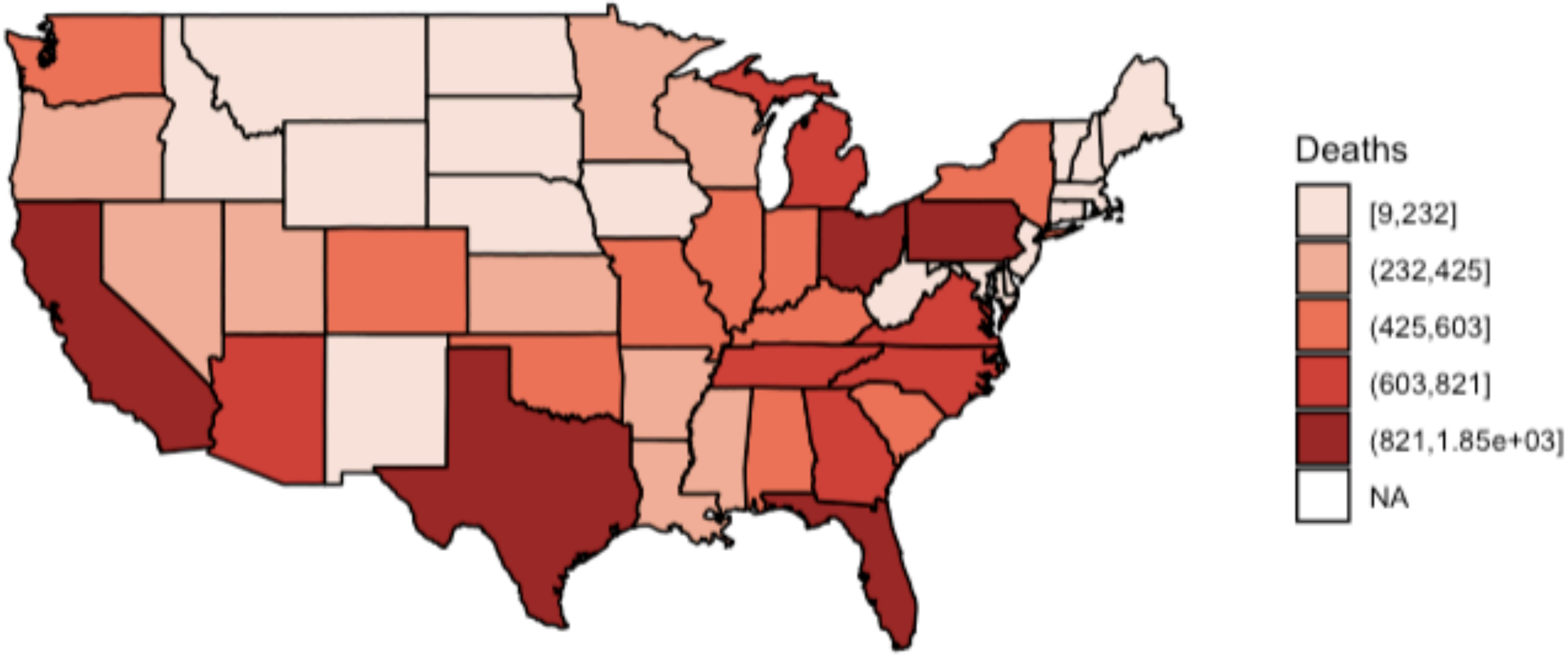
Gun Violence Archive

- Non-profit compiled dataset using media reports and publicly-available police reports
- Covers all 50 states 2013-present (updates real-time)
- Provides raw shooting-level data
- Includes fatal, non-fatal, and brandishing incidents

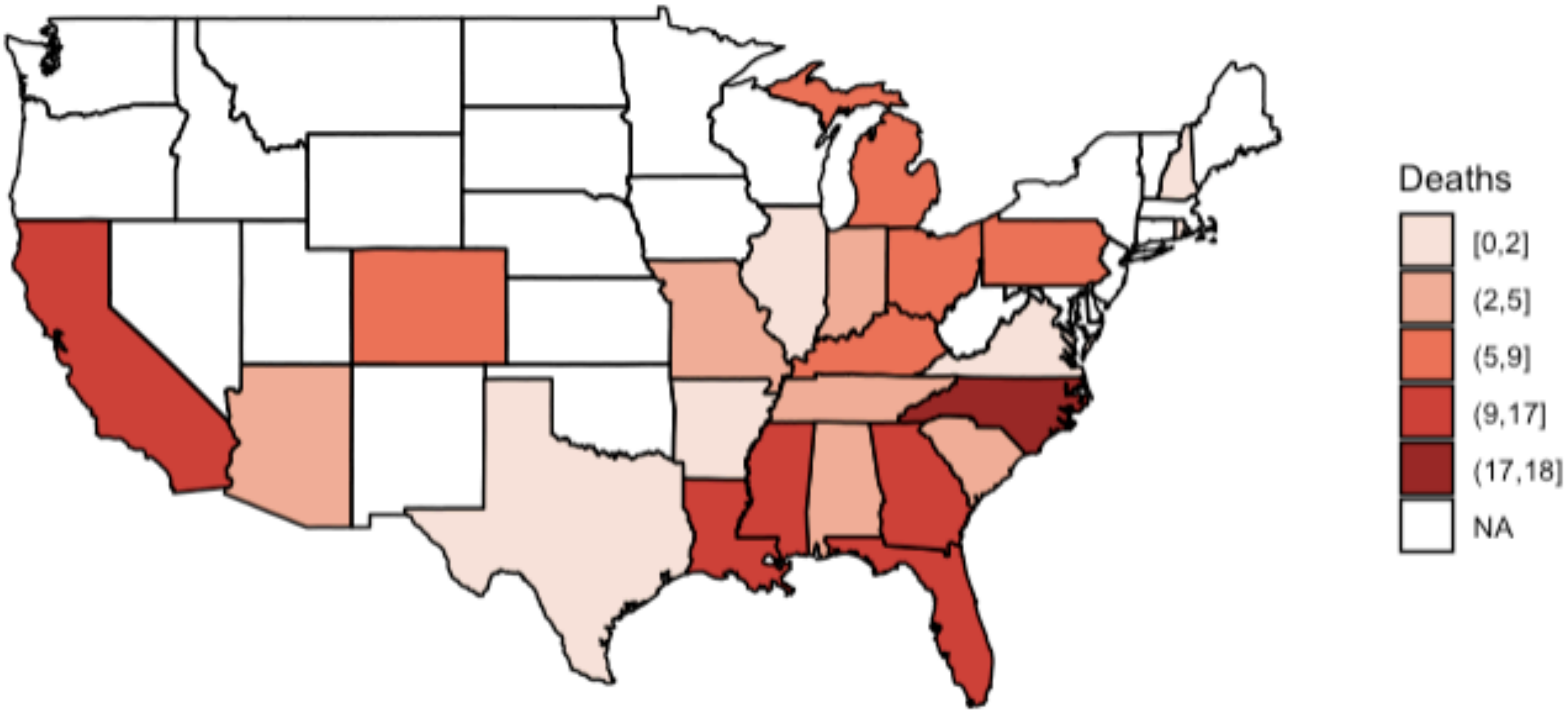
Range of Firearm Homicide Count Across NVDRS, NVSS, GVA, and FBI, 2016



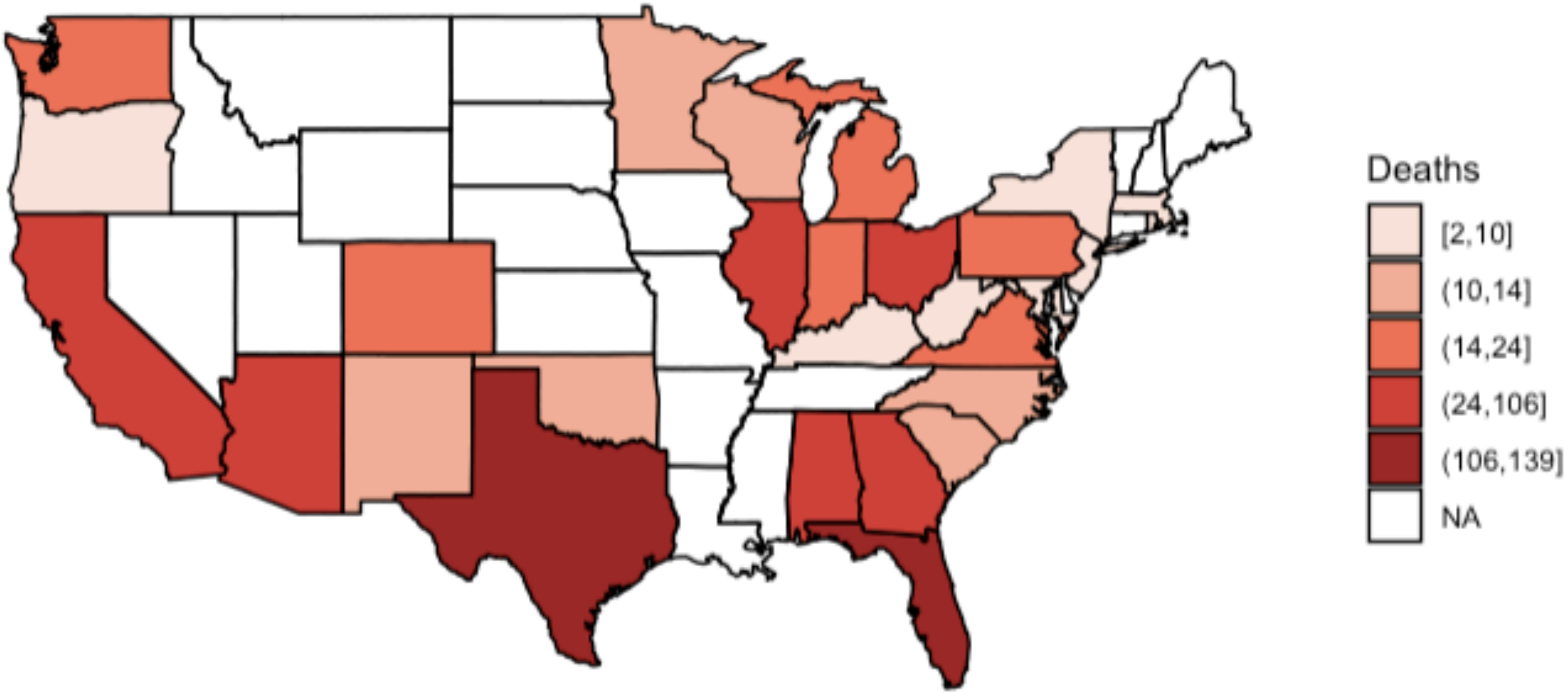
Range of Firearm Suicide Count Across NVDRS, NVSS, and GVA, 2016



Range of Firearm Accident Count Across NVDRS, NVSS, and GVA, 2016



Range of Firearm Legal Intervention Count Across NVDRS, NVSS, and GVA, 2016



Shooting Type	National Range, 2016
Homicide	3,415
Suicide	945
Accidental	153
Legal	798
Total	5,311

Data Source	Rate Ratio of Access
NVSS	62.8
NVDRS	12.55
FBI	58.56

Problem

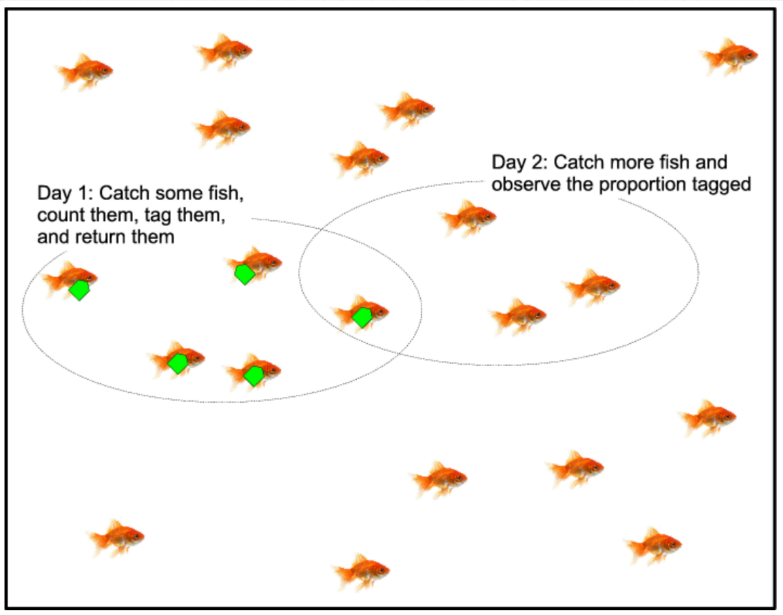
Can we use data integration techniques to combine the 3 federal data sources (NVDRS, NVSS, FBI) with the largest non-profit source (GVA) to obtain estimates of the true number of fatal shootings in the U.S. in 2016?

Challenges

1. Small cell suppression
2. Varying levels of summary data (individual level vs. counts)
3. Data source that is most likely to be unbiased (NVDRS) is missing almost half of U.S. states
4. Unknown overlap between data sources

Proposed Approach

- Following the approach of Royle (Biometrics 2009), we fit a Bayesian capture-recapture model with data augmentation


Day 1: Catch some fish, count them, tag them, and return them
Day 2: Catch more fish and observe the proportion tagged

Methods: Model

- Let the true number of shootings be N, and let it be sampled T times (T = 4, for our 4 datasets)
- This sampling yields n unique observations, each captured y_i times, $1 \le y_i \le 4$, $i = 1,\ldots, n$
- We can model detection probability, p_i, as function of covariates (race, sex, shooting intent, state)
- However, our sample is biased: **p_i is higher in the sample than it is in the total population**

Methods: Model

- To deal with this bias, introduce (M-n) augmentation rows with $y_i = 0$, for a total of M rows
- These zero-augmented rows have all their covariates missing
- Introduce a latent variable $z_i \sim \text{Bern}(\Psi)$: Ψ is the probability that the augmented data is part of the true population.
  - $z_i = 1$ in rows 1, …, n and is missing in rows n+1, …, M

Methods: Model

$$logit(p_i) = \beta_0 + \beta_1 sex_i + \beta_2 race_i + \beta_3 revenue_i + \beta_4 income_i + \beta_5 gun_i + \beta_6 intent_i,\ i = 1,...,M$$

$$y_i | p_i \sim Bin(T,\ p_i z_i)$$

$$z_i \sim Bern(\Psi),\ \Psi \sim U(0,1)$$

$$Sex_i \sim Bern(\pi),\ \pi \sim U(0,1)$$

$$Race_i,\ Intent_i \sim Multi(\gamma_i)$$

$$Revenue_i,\ Income_i \sim N(\mu, \sigma^2),\ \mu \sim N(0,1000),\ \sigma^{-2} \sim Gam(0.001,0.001)$$

$$Gun_i \sim N(\mu, \sigma^2),\ \mu \sim N(0,1),\ \sigma^{-2} \sim Gam(0.001,0.001)$$

$$\beta_i \sim N(0,1000)$$

Methods: Data Merge

- Biggest violated assumption of above approach: we know how to match the shootings across the datasets.
- Match the 4 datasets using as much available information as possible.
- Record the number of times each (hopefully unique) shooting was captured, and by which datasets.

Methods: Data Merge

Starting Data

State	Intent	Race	Sex	GVA
Alabama	Homicide	Black	Female	1
Alabama	Homicide	White	Male	1
Wyoming	Legal	NA	Male	1

External NVDRS Data

State	Intent	Race	Sex	Deaths
Alabama	Homicide	Black	Female	2
Alabama	Homicide	White	Male	0
Wyoming	Legal	White	Male	2

Methods: Data Merge

State	Intent	Race	Sex	GVA	NVDRS
Alabama	Homicide	Black	Female	1	1
Alabama	Homicide	Black	Female	0	1
Alabama	Homicide	White	Male	1	0
Wyoming	Legal	White	Male	1	1
Wyoming	Legal	White	Male	0	1

Methods: Data Merge

- NVSS estimated total number of fatal shootings to be 38,658
- After the merge, we estimate the number of unique shootings to be 41,682

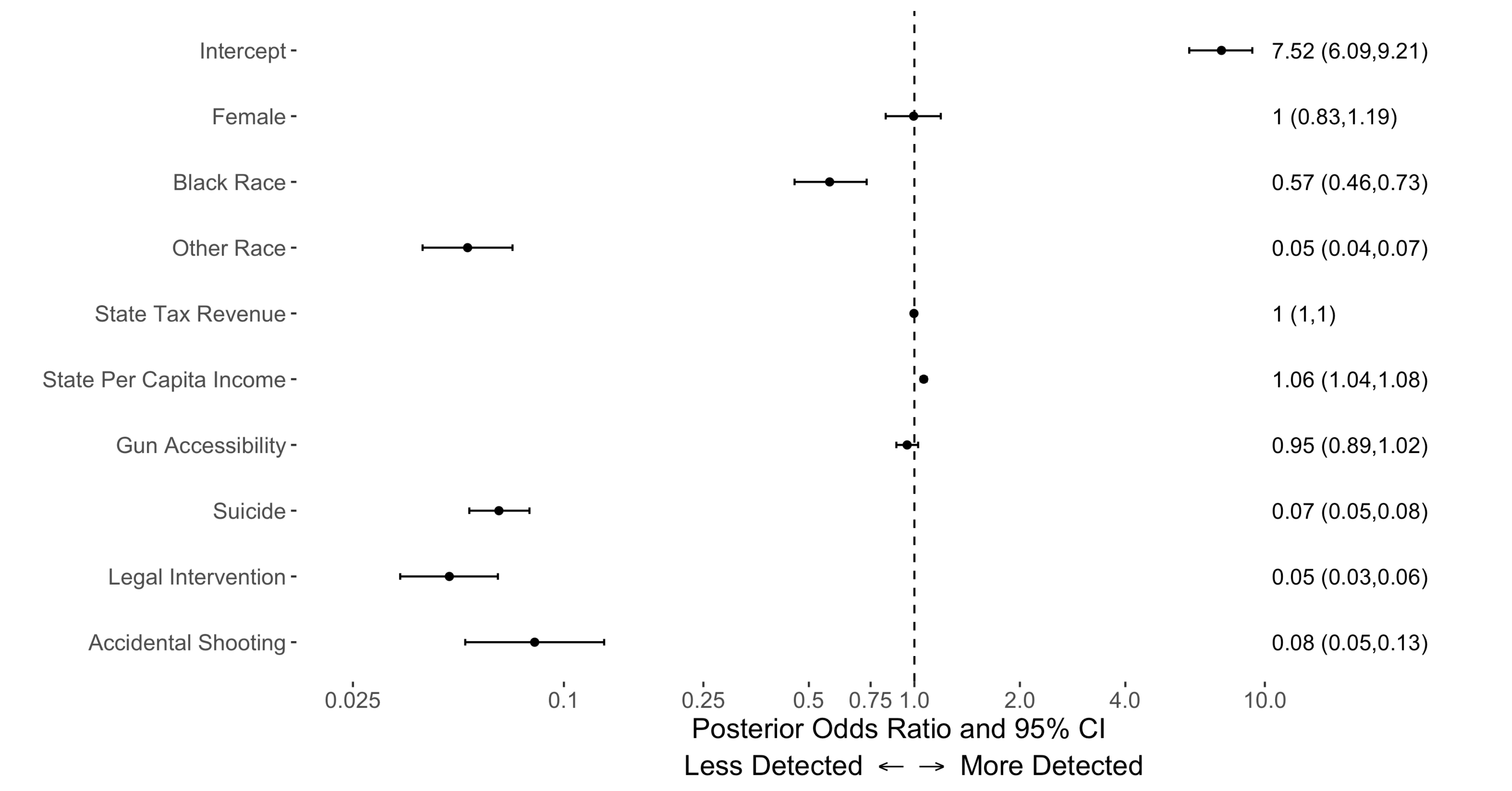
Results

- Implemented in JAGS
- Observed 41,682 shootings in 2016; augmented that with 10,000 zero-rows
- Due to computing constraints, used 5% sample of augmented data, so 2,584 shootings considered
- Ran MCMC for 21,000 iterations over 4 chains; discarded first 11,000 and thinned to every 5th
- At least 700 effective samples for every parameter; convergence looked good

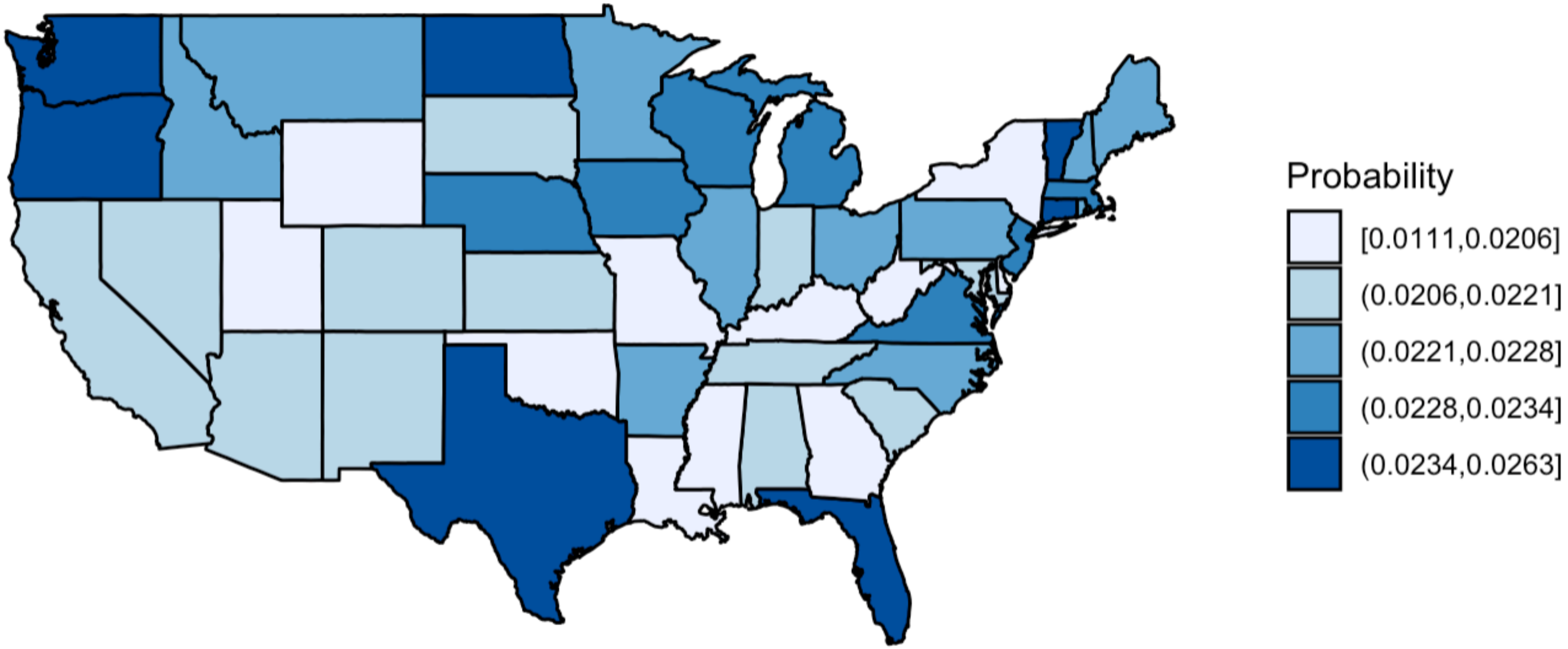
Estimated Shootings

- Model estimated total number of shootings in 2016 to be: 51,600 (51380, 51680)
- Observed shootings (all data): 41,682
- Compare to NVSS count: 38,658

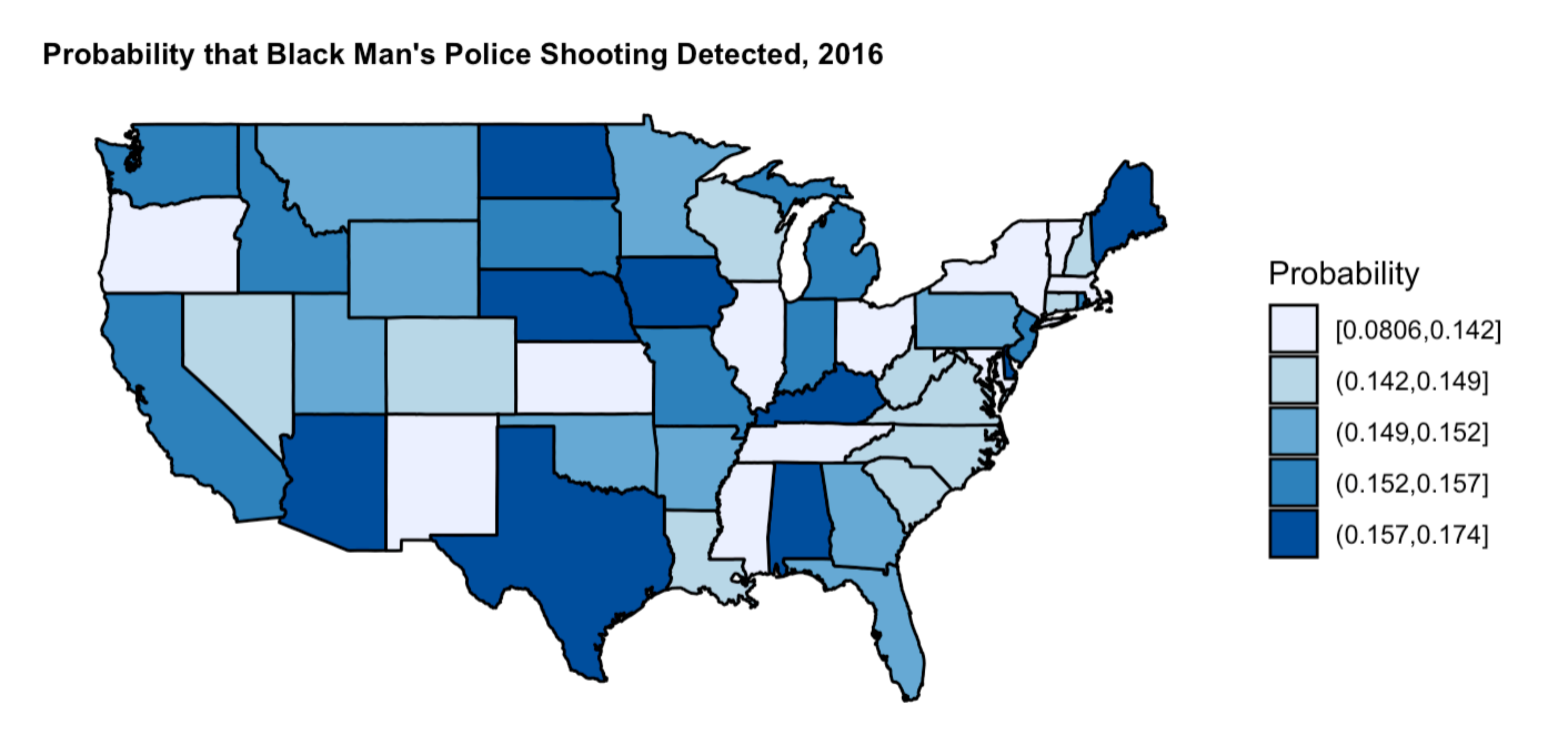
Insights into Detection



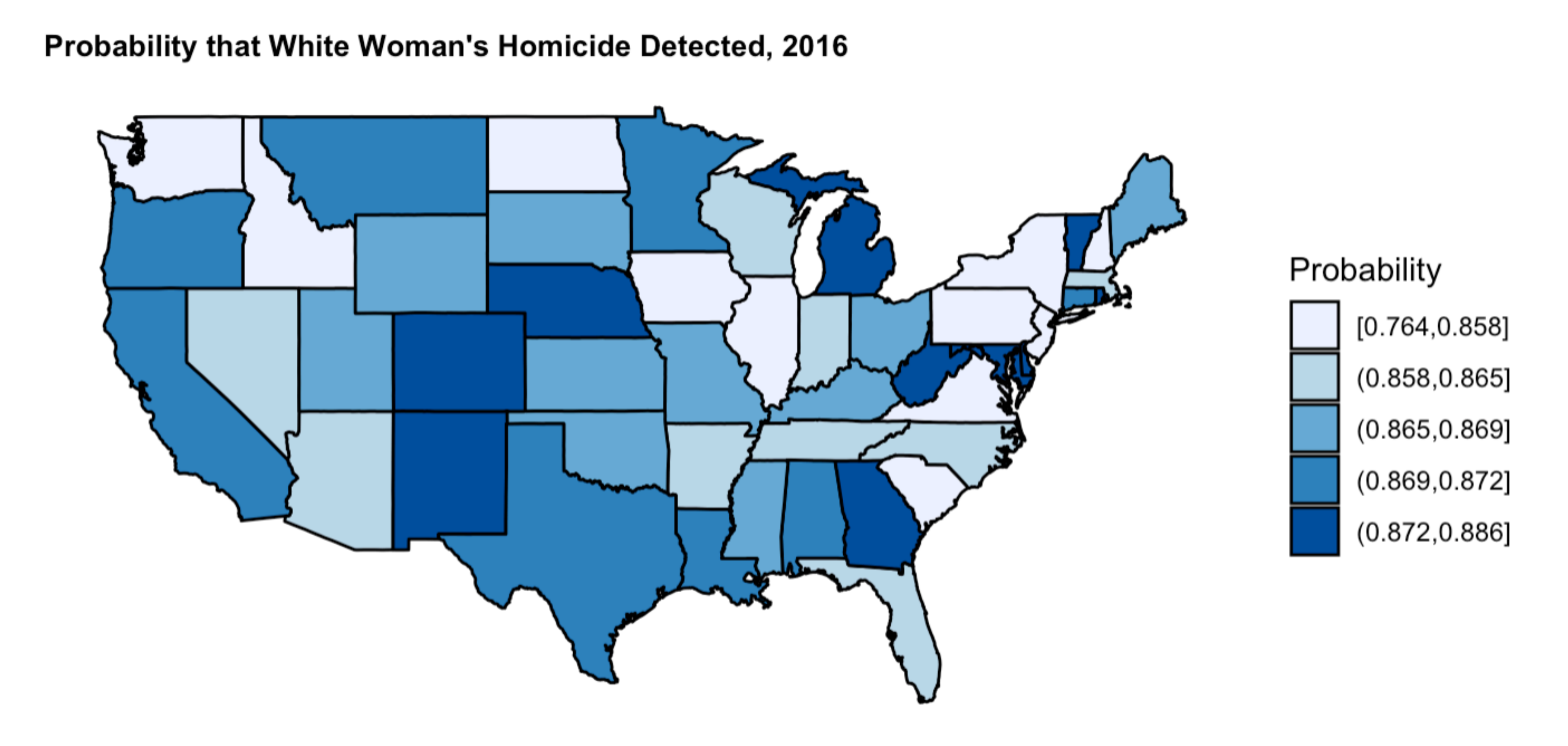
Insights into Detection


Probability that Other Race Man's Suicide Detected, 2016

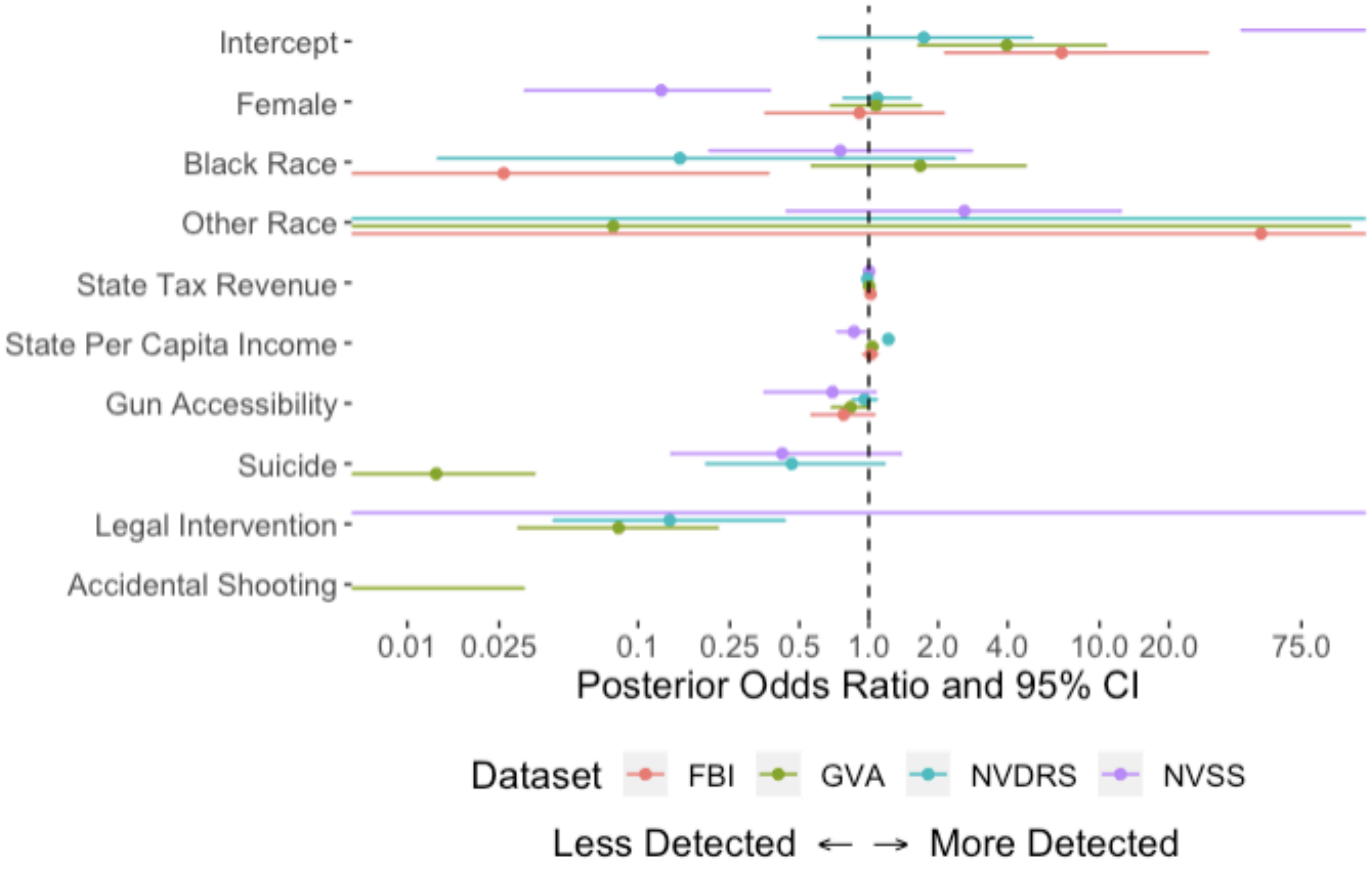
Insights into Detection



Insights into Detection



Stratified Model



Limitations & Future Work

- More nuanced Bayesian imputation to address suppressed cell counts
- Alternative approaches to case-matching
- Allow for varying effects of race and intent on probability of capture for each dataset; consider clustering of data
- Simulation studies and sensitivity analyses for priors

References

1. John W. Ayers, Benjamin M. Althouse, Eric C. Leas, Ted Alcorn, Mark Dredze. "Can Big Media Data Revolutionize Gun Violence Prevention?" Bloomberg Data for Good Exchange Conference, New York City, Sept. 25, 2016.
2. Mayors Against Illegal Guns, "Access Denied: How the Gun Lobby Is Depriving Police, Policy Makers, and the Public of the Data We Need to Prevent Gun Violence," January 2013.
3. Gun Violence Archive. https://www.gunviolencearchive.org/methodology, 2019.
4. CDC, National Violent Death Reporting System, 2019.
5. CDC, Fatal and Non-Fatal Injury in the United States, 2019.
6. FBI, Supplemental Homicide Reports, 2019.
7. Ruth King and Rachel McCrea, "Capture-Recapture Methods and Models: Estimating Population Size," in Integrated Population Biology and Modeling, Part 2, 2019.
8. J. Andrew Royle, "Analysis of Capture-Recapture Models with Individual Covariates Using Data Augmentation," Biometrics 65 (March 2009): 267-274.

Thank you! Questions?