Applied statistic 2

Replication of **Are Police Racially Biased in the Decision to Shoot?**

Clark, T.S. *et al.* (2023) "Are police racially biased in the decision to shoot?," *The Journal of Politics* [Preprint]. Darragh Kane O Toole

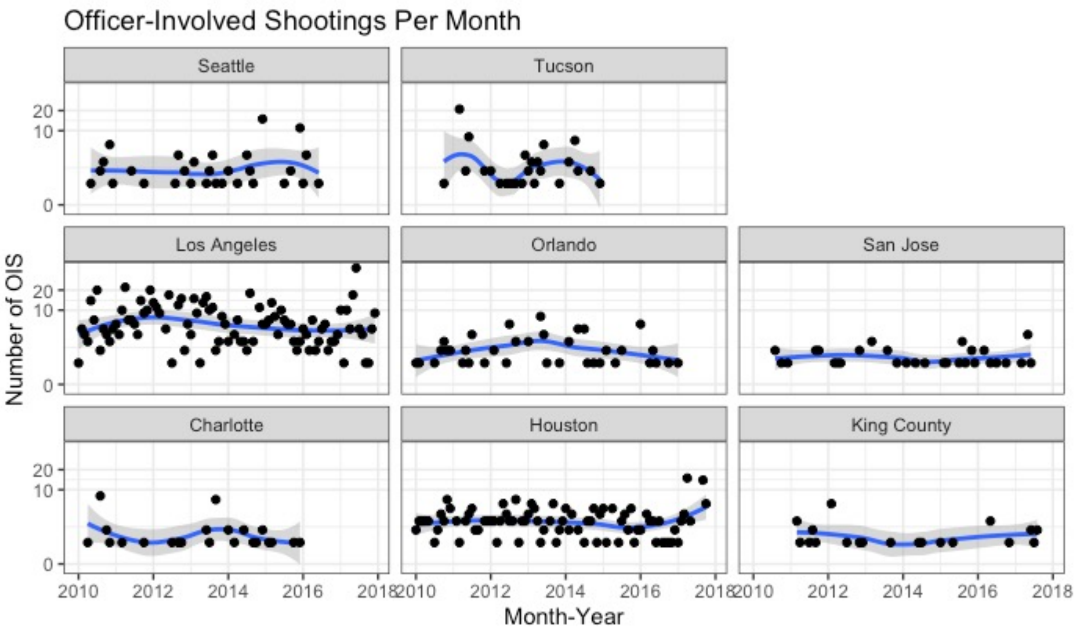
Abstract

- We present a theoretical model predicting racially biased policing produces 1) more use of potentially lethal force by firearms against Black civilians than against White civilians and 2) lower fatality rates for Black civilians than White civilians. We empirically evaluate this second prediction with original officer-involved shooting data from nine local police jurisdictions from 2005 to 2017, finding that Black fatality rates are significantly lower than White fatality rates, conditional upon civilians being shot by the police. Using outcome test methodology, we estimate that at least 30% of Black civilians shot by the police would not have been shot had they been White. We also show that an omitted covariate three times stronger than our strongest included covariate would only reduce this estimate to 18%. Additionally, such an omitted covariate would have to affect Black fatality rates and not Hispanic fatality rates in order to be consistent with the data.

Clark, T.S. *et al.* (2023) "Are police racially biased in the decision to shoot?," *The Journal of Politics* [Preprint].

- The research shows that the effect of civilian race has on the likelihood of a police officer shooting a civilian.
- "Our logic implies that if officers have a lower threshold for deciding to shoot Black civilians than White civilians, then there will be a greater proportion of Black civilians who will choose to not threaten and, therefore, survive an officer-involved shooting"

Officer shootings per month



This is the percent of fatal shootings based on civilian race

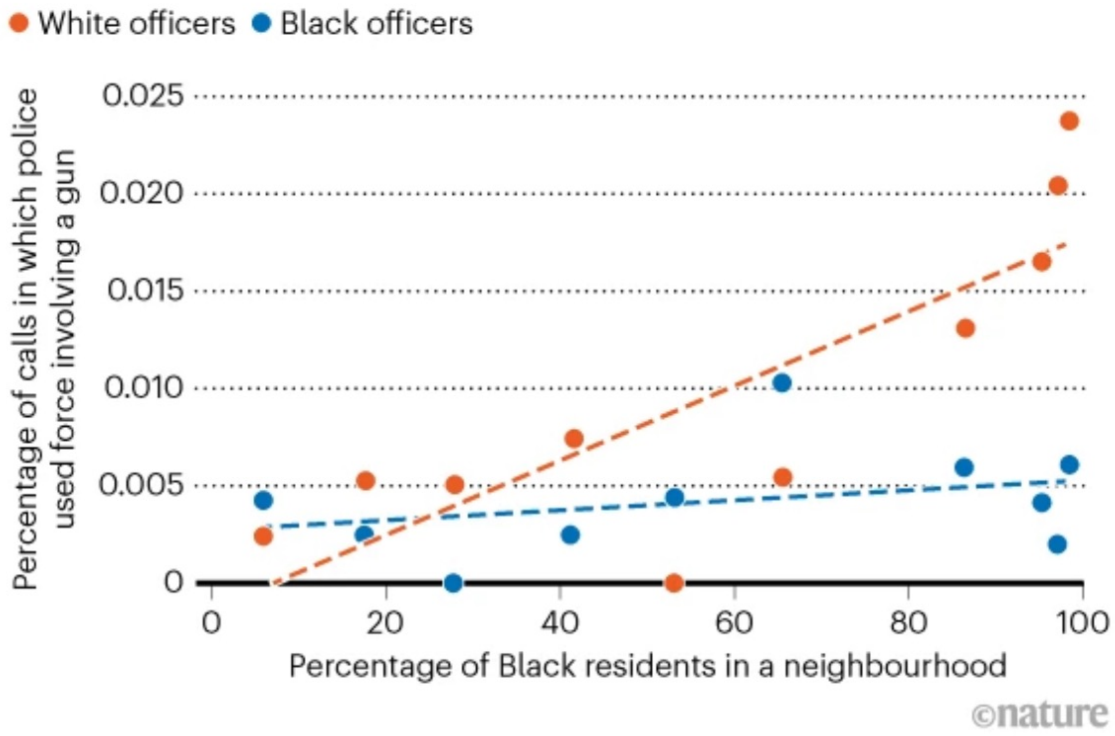
	White	Black	Hispanic	Asian/AI/AN/PI
0	0.48	0.67	0.42	0.27
1	0.52	0.33	0.58	0.73

The original vs replication

- The dataset includes a wide range if data but the studies focus is on
- fatal or not
- Race-Asian,Black,White or Hispanic
- The city Charlotte, Houston ,King County,Los Angeles,Orlando,San Jose,Seattle,Tucson
- Distance from a trauma center(hospital)

- Replication
- In my replication I ran the model, but I added a new variable, the race of the police officer to the fixed effect additive model

Officer Race and shooting
Does it matter?



Yes-It does in a lot of datasets including this one from 2020

(Hoekstra & Sloan, 2020)

Replication twist-Surely officer race matters?

I created a new alternative dataset that included the race variable for police.

I then adjusted the functions made by the authors to enable this change to run

To align with methodology used where they ran 4 models each with an extra variable, I did similarly adding a 5th model which included this new variable and reran all new code

Fixed effect glm for original and replication

```r
# Table 2 main analysis Logistic -------
glm_main1 <- feglm(fatal~Black + Hispanic + Asian,
                   data=orig,
                   family=binomial(link="logit"),
                   se="cluster",
                   cluster="new_group_id")

glm_main2 <- feglm(fatal~Black + Hispanic + Asian +city_clean,
                   data=orig,
                   family=binomial(link="logit"),
                   se="cluster",
                   cluster="new_group_id")
glm_main3 <- feglm(fatal~Black + Hispanic + Asian +trauma10,
                   data=orig,
                   family=binomial(link="logit"),
                   se="cluster",
                   cluster="new_group_id")
glm_main4 <- feglm(fatal~Black + Hispanic + Asian +trauma10 +
                     city_clean,
                   data=orig,
                   family=binomial(link="logit"),
                   se="cluster",
                   cluster="new_group_id")

```

```r
                   data=origo,
                   family=binomial(link="logit"),
                   se="cluster",
                   cluster="new_group_id")
glmo2 <- feglm(fatal~Black + Hispanic + Asian +city_clean ,
                   data=origo,
                   family=binomial(link="logit"),
                   se="cluster",
                   cluster="new_group_id")
glmo3 <- feglm(fatal~Black + Hispanic + Asian+ trauma10   ,
                   data=origo,
                   family=binomial(link="logit"),
                   se="cluster",
                   cluster="new_group_id")
glmo4 <- feglm(fatal~Black + Hispanic + Asian +trauma10 +
                     city_clean,
                   data=origo,
                   family=binomial(link="logit"),
                   se="cluster",
                   cluster="new_group_id")

glmo5 <- feglm(fatal~Black + Hispanic + Asian +trauma10 +
                 city_clean+ officer_race_new,
               data=origo,
               family=binomial(link="logit"),
               se="cluster",
               cluster="new_group_id")
```

Adapting sensitivity function

- The estimate.poisson.rhoU2 function is estimating a Poisson regression model with log link and using the confounder variable as the predictor of interest, with White, Houston, King_County, Los_Angeles, Orlando, San_Jose, Seattle, Tucson, and trauma10 as covariates. The function returns the fitted Poisson regression model.
- The bias.factor function calculates the bias factor based on the ratio of the unmeasured confounding effect (represented by RR_rhoU) and the observed exposure effect (represented by RR_UY).
- The cluster.pois2 function appears to be fitting a clustered Poisson regression model using the fepois function, with fatal, White, trauma10, and city_factor as predictors. The function returns the estimated coefficients of the fitted model. The indices argument likely refers to the indices of the observations in the data set that belong to the same cluster.
-

```r
# with intercept, Charlotte as ref
estimate.poisson.rhoU2 <- function(data,confounder){
  vars <- c("White","Houston","King_County","Los_Angeles",
            "Orlando","San_Jose","Seattle",
            "Tucson","trauma10")
  vars_rhs <- vars[!vars %in% confounder]
  f <- as.formula(paste(confounder,
                        paste(vars_rhs,collapse = "+"),
                        sep="~"))
  fit <- glm(f, data = data,
             family = poisson(link = "log"))
  return(fit)
}

bias.factor <- function(RR_UY, RR_rhoU){
  b <- (RR_UY*RR_rhoU) / (RR_UY + RR_rhoU - 1)
  return(b)
}

cluster.pois2 <- function(data, indices){
  d <- data[indices,]
  fit <- fepois(fatal ~ White + trauma10 + city_factor,
                data = d,
                se = "cluster",
                cluster = cluster_variable)
  coef <- fit$coefficients
```

```r
# with intercept, Charlotte as ref
estimate.poisson.rhoU2 <- function(data,confounder){
  vars <- c("White","Houston","King_County","Los_Angeles",
            "Orlando","San_Jose","Seattle",
            "Tucson","trauma10","officer_race_new")
  vars_rhs <- vars[!vars %in% confounder]
  f <- as.formula(paste(confounder,
                        paste(vars_rhs,collapse = "+"),
                        sep="~"))
  fit <- glm(f, data = data,
             family = poisson(link = "log"))
  return(fit)
}

bias.factor <- function(RR_UY, RR_rhoU){
  b <- (RR_UY*RR_rhoU) / (RR_UY + RR_rhoU - 1)
  return(b)
}

cluster.pois2 <- function(data, indices){
  d <- data[indices,]
  fit <- fepois(fatal ~ White + trauma10 + city_factor + officer_race_
                data = d,
                se = "cluster",
                cluster = cluster_variable)
  coef <- fit$coefficients
  return(coef)
}
```

Estimated relationship between civilian race and probability of fatality conditional upon being involved in an officer-involved shooting. Cells show logistic coefficients with clustered-robust standard errors. Omitted category is white civilians, Charlotte and the year 2010."

A excerpt of the code used to generate these statistics

```r
glm_main4 <- feglm(fatal~Black + Hispanic + Asian + trauma10 +
                      city_clean+year_factor,
                    data=orig,
                    family=binomial(link="logit"),
                    se="cluster",
                    cluster=cluster_variable)

# name mapping list
coefsnames <- list('Black' = 'Black',
                    "Hispanic" = 'Hispanic',
                    'Asian' = 'Asian/AI/AN/PI',
                    'trauma10' = 'Closest Trauma (10s miles)',
                    'city_cleanHouston' = 'Houston',
                    'city_cleanKing County' = 'King County',
                    'city_cleanLos Angeles' = 'Los Angeles',
                    'city_cleanOrlando' = 'Orlando',
                    'city_cleanSan Jose' = 'San Jose',
                    'city_cleanSeattle' = 'Seattle',
                    'city_cleanTucson' = 'Tucson',
                    'year_factor2011' = '2011',
                    'year_factor2012' = '2012',
                    'year_factor2013' = '2013',
                    'year_factor2014' = '2014',
                    'year_factor2015' = '2015',
                    'year_factor2016' = '2016',
                    'year_factor2017' = '2017',
                    '(Intercept)' = 'Intercept')

screenreg(list(glm_main1,glm_main2,glm_main3,glm_main4),
          custom.coef.map = coefsnames,
          omit.coef = ":",
          include.dev = FALSE,
          include.loglik = FALSE,
          include.pseudors = FALSE)
```

	Model 1	Model 2	Model 3	Model 4
Black	-0.67*	-0.82**	-0.74*	-0.74*
	(0.33)	(0.31)	(0.33)	(0.33)
Hispanic	0.13	0.28	0.11	0.11
	(0.31)	(0.31)	(0.32)	(0.32)
Asian/AI/AN/PI	0.91	1.04	0.96	0.96
	(0.58)	(0.64)	(0.60)	(0.60)
Closest Trauma (10s miles)	0.41	0.19	0.45*	0.45*
	(0.21)	(0.18)	(0.21)	(0.21)
Houston	-0.21		-0.19	-0.19
	(0.63)		(0.59)	(0.59)
King County	-0.12		-0.17	-0.17
	(0.81)		(0.79)	(0.79)
Los Angeles	1.15		1.18*	1.18*
	(0.62)		(0.58)	(0.58)
Orlando	0.52		0.50	0.50
	(0.70)		(0.67)	(0.67)
San Jose	0.16		0.18	0.18
	(0.69)		(0.67)	(0.67)
Seattle	1.27		1.20	1.20
	(0.77)		(0.76)	(0.76)
Tucson	1.64*		1.54*	1.54*
	(0.73)		(0.69)	(0.69)
2011		0.51	0.33	0.33
		(0.41)	(0.40)	(0.40)
2012		0.25	0.33	0.33
		(0.43)	(0.42)	(0.42)
2013		0.34	0.30	0.30
		(0.43)	(0.42)	(0.42)
2014		0.63	0.63	0.63
		(0.47)	(0.45)	(0.45)
2015		0.20	0.27	0.27
		(0.44)	(0.42)	(0.42)
2016		0.13	0.28	0.28
		(0.44)	(0.41)	(0.41)
2017		-0.13	-0.09	-0.09
		(0.52)	(0.50)	(0.50)
Intercept	-1.00	-0.31	-1.27	-1.27
	(0.64)	(0.41)	(0.66)	(0.66)
Num. obs.	1250	1250	1250	1250

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

	Model 1	Model 2	Model 3	Model 4	Model 5
Black	-0.84**	-0.77*	-0.82**	-0.74*	-0.70*
	(0.30)	(0.33)	(0.31)	(0.33)	(0.35)
Hispanic	0.27	0.08	0.28	0.11	0.16
	(0.31)	(0.33)	(0.31)	(0.32)	(0.34)
Asian/AI/AN/PI	1.00	0.84	1.04	0.96	1.06
	(0.63)	(0.59)	(0.64)	(0.60)	(0.61)
Closest Trauma (10s miles)			0.19	0.45*	0.47*
			(0.18)	(0.21)	(0.22)
Houston		0.06		-0.19	-0.16
		(0.56)		(0.59)	(0.59)
King County		0.30		-0.17	-0.02
		(0.75)		(0.79)	(0.81)
Los Angeles		1.31*		1.18*	1.16
		(0.56)		(0.58)	(0.59)
Orlando		0.58		0.50	0.53
		(0.64)		(0.67)	(0.68)
San Jose		0.22		0.18	
		(0.65)		(0.67)	
Seattle		1.18		1.20	1.20
		(0.74)		(0.76)	(0.77)
Tucson		1.54*		1.54*	1.52*
		(0.67)		(0.69)	(0.70)
2011	0.45	0.19	0.51	0.33	0.34
	(0.41)	(0.41)	(0.41)	(0.40)	(0.41)
2012	0.20	0.14	0.25	0.33	0.33
	(0.43)	(0.43)	(0.43)	(0.42)	(0.43)
2013	0.32	0.17	0.34	0.30	0.31
	(0.42)	(0.41)	(0.43)	(0.42)	(0.43)
2014	0.57	0.44	0.63	0.63	0.64
	(0.47)	(0.45)	(0.47)	(0.45)	(0.47)
2015	0.17	0.13	0.20	0.27	0.21
	(0.44)	(0.42)	(0.44)	(0.42)	(0.45)
2016	0.14	0.14	0.13	0.28	0.24
	(0.43)	(0.41)	(0.44)	(0.41)	(0.42)
2017	-0.14	-0.24	-0.13	-0.09	-0.15
	(0.52)	(0.51)	(0.52)	(0.50)	(0.51)
Intercept	-0.14	-0.92	-0.31	-1.27	1.32*

```r
lpm_main4 <- feols(fatal~Black + Hispanic + Asian + trauma10 +
                         city_clean*year_factor,
                       data=orig,
                       se="cluster",
                       cluster=cluster_variable)
capt <- "Estimated relationship between civilian race and probability of
screenreg(list(lpm_main1,lpm_main2,lpm_main3,lpm_main4),
          omit.coef = "city|year")
# save table
texreg(list(lpm_main1,lpm_main2,lpm_main3,lpm_main4),
       custom.coef.map = coefsnames,
       omit.coef = ":",
       include.dev = FALSE,
       include.loglik = FALSE,
       include.pseudors = FALSE,
       include.rs = FALSE,
       include.adjrs = FALSE,
       caption = capt,
       booktabs = TRUE,
       use.package=FALSE,
       float.pos = "ht!",
```

- "Estimated relationship between civilian race and probability of fatality conditional upon being involved in an officer-involved shooting. Cells show linear probability model coefficients with clustered-robust standard errors. Omitted category is white civilians, Charlotte and the year 2010."

(Intercept)	0.29 *	0.43 ***	0.23	-0.06 *
	(0.12)	(0.10)	(0.13)	(0.02)
Black	-0.15 *	-0.19 **	-0.17 *	-0.16 *
	(0.07)	(0.07)	(0.07)	(0.07)
Hispanic	0.03	0.07	0.03	0.03
	(0.07)	(0.07)	(0.07)	(0.07)
Asian	0.19	0.24	0.20	0.18
	(0.11)	(0.13)	(0.11)	(0.13)
trauma10	0.09 *	0.05	0.10 *	0.10 *
	(0.04)	(0.04)	(0.04)	(0.04)
Num. obs.	1250	1250	1250	1250
R^2 (full model)	0.13	0.07	0.14	0.19
R^2 (proj model)				
Adj. R^2 (full model)	0.12	0.06	0.12	0.15
Adj. R^2 (proj model)				

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	0.29 *	0.43 ***	0.23	-0.06 *	-0.07 *
	(0.12)	(0.10)	(0.13)	(0.02)	(0.03)
Black	-0.15 *	-0.19 **	-0.17 *	-0.16 *	-0.14
	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Hispanic	0.03	0.07	0.03	0.03	0.04
	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Asian	0.19	0.24	0.20	0.18	0.21
	(0.11)	(0.13)	(0.11)	(0.13)	(0.12)
trauma10	0.09 *	0.05	0.10 *	0.10 *	0.11 **
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
officer_race_newBlack					-0.04
					(0.05)
officer_race_newHispanic					0.03
					(0.04)
officer_race_newAsian/AI/AN/PI					0.07
					(0.06)
Num. obs.	1250	1250	1250	1250	1193
R^2 (full model)	0.13	0.07	0.14	0.19	0.19
R^2 (proj model)					
Adj. R^2 (full model)	0.12	0.06	0.12	0.15	0.15
Adj. R^2 (proj model)					

Estimated relationship between civilian race and probability of fatality conditional upon being involved in an officer-involved shooting. Cells show Poisson coefficients with clustered-robust standard errors. Omitted category is white civilians, Charlotte and the year 2010."

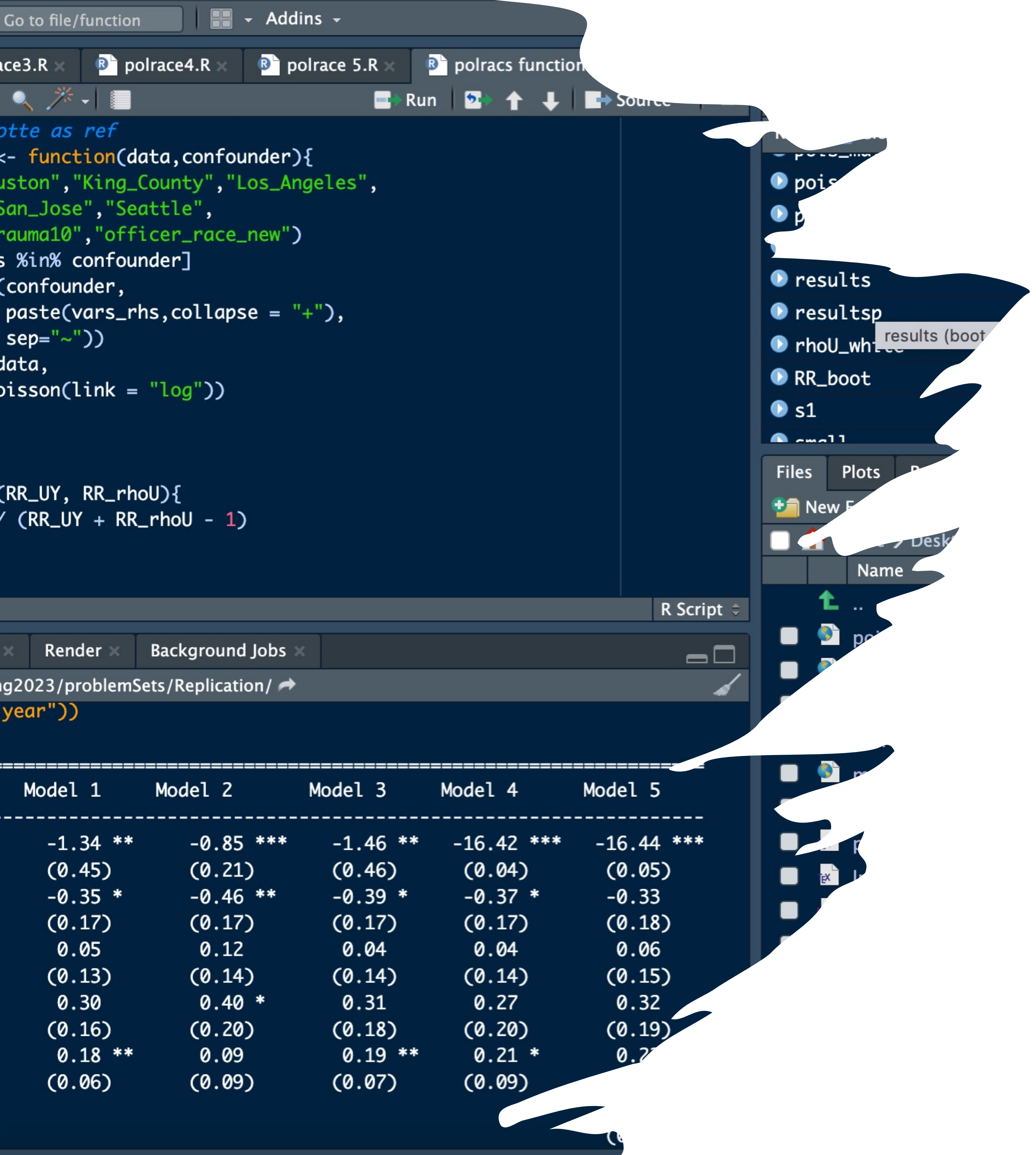
```r

pois_main3 <- fepois(fatal~Black + Hispanic + Asian + trauma10 +
                         city_clean + year_factor,
                       data=orig,
                       se="cluster",
                       cluster=cluster_variable)


pois_main4 <- fepois(fatal~Black + Hispanic + Asian + trauma10 +
                         city_clean*year_factor,
                       data=orig,
                       se="cluster",
                       cluster=cluster_variable)
capt <- "Estimated relationship between civilian race and probability of fatality
screenreg(list(pois_main1,pois_main2,pois_main3,pois_main4),
          omit.coef=c("city|year"))
# save table
texreg(list(pois_main1,pois_main2,pois_main3,pois_main4),
       custom.coef.map = coefsnames,
       omit.coef = ":",
       include.dev = FALSE,
       include.loglik = FALSE,
       include.pseudors = FALSE,
       caption = capt,
       booktabs = TRUE,
       use.package=FALSE,
       float.pos = "ht!",
       file = here("JOP_racialbias/tables","poisson-FE-models.tex"),
       label="tab:poisson_FE")
```

	Model 1	Model 2	Model 3	Model 4
(Intercept)	-1.34 **	-0.85 ***	-1.46 **	-16.42 ***
	(0.45)	(0.21)	(0.46)	(0.04)
Black	-0.35 *	-0.46 **	-0.39 *	-0.37 *
	(0.17)	(0.17)	(0.17)	(0.17)
Hispanic	0.05	0.12	0.04	0.04
	(0.13)	(0.14)	(0.14)	(0.14)
Asian	0.30	0.40 *	0.31	0.27
	(0.16)	(0.20)	(0.18)	(0.20)
trauma10	0.18 **	0.09	0.19 **	0.21 *
	(0.06)	(0.09)	(0.07)	(0.09)
Num. obs.	1250	1250	1250	1250
Deviance	790.26	832.79	785.46	739.62
Log Likelihood	-997.13	-1018.39	-994.73	-971.81
Pseudo R^2	0.03	0.01	0.03	0.01

*** p < 0.001; ** p < 0.01; * p < 0.05

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	-1.34 **	-0.85 ***	-1.46 **	-16.42 ***	-16.44 ***
	(0.45)	(0.21)	(0.46)	(0.04)	(0.05)
Black	-0.35 *	-0.46 **	-0.39 *	-0.37 *	-0.33
	(0.17)	(0.17)	(0.17)	(0.17)	(0.18)
Hispanic	0.05	0.12	0.04	0.04	0.06
	(0.13)	(0.14)	(0.14)	(0.14)	(0.15)
Asian	0.30	0.40 *	0.31	0.27	0.32
	(0.16)	(0.20)	(0.18)	(0.20)	(0.19)
trauma10	0.18 **	0.09	0.19 **	0.21 *	0.22 **
	(0.06)	(0.09)	(0.07)	(0.09)	(0.08)
officer_race_newBlack					-0.12
					(0.15)
officer_race_newHispanic					0.05
					(0.07)
officer_race_newAsian/AI/AN/PI					0.14
					(0.10)
Num. obs.	1250	1250	1250	1250	1193
Deviance	790.26	832.79	785.46	739.62	702.33
Log Likelihood	-997.13	-1018.39	-994.73	-971.81	-936.17
Pseudo R^2	0.03	0.01	0.03	0.01	0.01

*** p < 0.001; ** p < 0.01; * p < 0.05



Code

- This is only a snippets of most relevant changes to code made
- To run my changes run I created 4 new code files polrace3,4,5 and polracs function each mirroring part 3,4,5 and the sensitivity function of the original code

Conclusion

The authors did not include the officer race and as all the statistics show they were right not to.

This dataset would indicate that the race of the officer is not statistically significant or have any notable impact on the dataset as far as I can prove.

Given the evidence for officer race not being a relevant factor and evidence from the study mentioned earlier I am left unsure of what to make of the officer race factor.

Thanks for listening

Any questions?