Assessing the impact of police officer line-of-duty killings and force size on civilian arrests

Srinidhi Narayanan

Abstract

The death of a police officer in the line of duty can be traumatic for the fallen officer's colleagues, and may elicit a response with downstream consequences — does the death of an officer in the line of duty affect how the force polices in the aftermath? This paper investigates this question by assessing the impact of an officer death on arrest numbers in the officer's municipality for up to six months post-death. Furthermore, the paper investigates the role of police force size in this impact. Both impacts are examined using a difference-in-differences approach, with a triple difference estimator being used in the second case. I find that neither death or an officer in the line of duty, nor police force size, have statistically significant effects on arrest numbers in the months considered; ignoring significance, the death of an officer has a negative impact on arrests in the post-treatment months, with a trend suggesting that months beyond 6 may experience significant effects.

1 Introduction

Policing is a key social and political issue throughout the world. This is even more the case in the United States, a country which incarcerates disproportionate to its size and in which, compared to other democracies, civilian-officer relationships are unusually strained [1] [2]. The state of policing in a country is an indicator of the nation's attitude toward crime and justice, notions of equality, and much more. It is therefore imperative that we develop a good understanding of policing mechanisms and the drivers of police behavior.

The death of a police officer in the line of duty is a shock that is oftentimes traumatic for the fallen officer's colleagues [8], and which may elicit a response in the form of changed policing behavior [4]. If an officer is killed in the line of duty, it is possible that other officers may engage in risk-mitigating behavior in the aftermath, being more reluctant to police or unwilling to interact with the public, possibly causing arrest numbers to decrease. On the other hand, they might become hyper-vigilant, prompting them to over-police, possibly driving arrest numbers upward. A body of work studying this question already exists (most notably [3]), but studies actually vary in their findings, even up to the direction of the impact.

It is also possible that the size of the affected police force is a contributor to this impact. In a smaller police force, the death of an officer may feel more personal to colleagues and may have an outsize impact on the change in arrest numbers; the size of the force may also indicate the level of support or resources the department experiences from the city, with a larger force benefitting from more support. Support and resources could reduce the impact of the officer death [5].

In summary, my paper will investigate the two-fold question "How does the death of an officer in the line of duty affect future policing behavior? And, does this effect vary based on police force size?" Though the former question has been answered before with the same data source I use in [3], I introduce a data sample restriction that I hypothesize will provide more accurate estimates of the coefficients derived in that paper. This paper also adds a previously-unexamined heterogeneity study with the investigation of the effect of police force size.

To this end, I examine FBI datasets with samples from 20,000 municipalities over a five year period from 2013-2017, and my empirical strategy uses a difference-in-differences design that exploits the staggered occurrence of line-of-duty deaths across municipalities and time. The initial analysis finds that arrests in municipalities where an officer is killed in the line of duty decrease in the months post-death though not by a statistically significant amount. With the addition of the force size as another difference, I then use a triple-difference estimator [9] to examine the impact of force size as well, also finding no significant effect.

The remainder of the paper will highlight the datasets and data preparation; describe the empirical framework and key assumptions of the difference-in-differences analysis; summarize results; and examine possible future steps.

2 Data

2.1 Datasets

I use FBI datasets, further cleaned/concatenated and available with more granularity via openICPSR [6] [7]. All data is available by municipality by month for the years 2013 to 2017 unless explicitly stated otherwise. I use data on officers assaulted and killed (LEOKA) and uniform crime reporting (UCR) for arrests/traffic stop numbers. I also make use of employment data that is contained in the LEOKA dataset.

Each dataset initially contains data on around 20,000 municipalities. LEOKA contains columns on the breakdowns of various types of assault; demographic information on the assailant(s) and victim(s); the populations of each municipality in each year; and as stated above, information on the numbers of male and female federal and civilian officers in each year. UCR contains columns pertaining to arrests of different types, weapons involved in those arrests, etc.

An initial overview of quantities of interest, prior to sample restrictions, is here:

Table 1: Summary Statistics

	Variables			
	Mean	S.D.	Min	Max
Population	33419	100625	0	4007905
Total Employee Officers	61.5	252.3	0	9988
Officers Killed	.0005	.03	0	5
Total Officers Assaulted	.62	4.05	0	566
Total Arrests	21.7	80.9	-55	20000

2.2 Sample Restriction

It is evident from the range of populations represented that municipalities vary significantly in size, and it is therefore reasonable to assume they would vary in many other ways as well. While adding place fixed effects to account for demographic and other differences is an option, a sample restriction could act as another control. I choose to restrict the data to only municipalities where an officer was at least shot, and consider the treatment to be an officer being killed. With this control, the officer's being killed becomes a quasi-random occurrence, and is thus a natural variation I make use of. After adding the sample restriction, the size of the dataset reduces by about 50 percent to around 10,000 municipalities.

2.3 Preprocessing

After retrieving the datasets from the openICPSR repositories, I apply some preprocessing in pandas before writing the data to stata .dta files for use in regressions. For example, I drop the unncessary last 200 columns from the LEOKA dataset to speed up calculations; add a total-assaults variable and an assault indicator variable; and encode the months from January of the first year as a 0 to December of the last year as a 60.

Since we are interested in the time trajectory of officer behavior, we want data for several months post treatment (a killing in a municipality in a month). We would also like to verify that there is no pre-treatment effect, so we want data for some months pre-treatment as well. So, I create six dummy variables for a killed officer 1 month prior, 2 months prior, . . . , 6 months prior; when an officer is killed in a municipality j in month t, the dummy column for 1 month prior is updated with a 1 in the row corresponding to municipality j, month $t + 1$; and so forth. Likewise, I create 1, 2, and 3 month prior pre-treatment dummies. The assumption in this creation is that the effect of the death, if any exists, will be apparent in six months, and that there is no treatment effect after six months.

I transform the arrests (dependent variable) column to a log of the arrests, since the interpretation of the regressions will then be with respect to percent changes; using raw arrest numbers could yield problems where results are skewed by some municipalities with very large or very small populations.

Finally, since we are interested in police force size, I calculate sizes, measured two ways: the first is a notion of "absolute" force size, and is a 1/0 indicator for large/small police force based on whether the force is larger or smaller than the median force size. I also calculate a median-thresholded size calculation based on the metric $\frac{\text{absolute force size}}{\text{population of municipality}}$.

The resultant data is a panel dataset with time variable being the numerically encoded month, and the place variable being the municipality.

3 Empirical Strategy

I employ a difference in differences strategy to this analysis, making some key assumptions in the process.

The most important is the parallel trends assumption, namely, that in the absence of a police killing, the arrest trends in untreated and treated places are similar. A previous paper that investigates the question of trends post police-killing [3] conducts several robustness and balance tests that appear to verify this assumption. In the results described below, it also becomes clear that the pre-treatment coefficients are not significantly different from zero, which is consistent with this assumption.

I assume that police force size doesn't change based on the arrest numbers (no reverse causation), even though in practice, it may be likely that more police are hired in response to rising arrest numbers since it seems like departments may need more manpower. As a simplification to the regression calculations, I use notions of size calculated with sizes from the first month of 2013. Thus, it is imperative that this assumption be close to true so that the assumed values of the variables are still valid in the later months of the study.

Related to the previous assumption, it is also important the municipality populations don't change in response to arrest numbers, since one view of force size incorporates the municipality population. In practice, it is also possible that place populations change in response to increasing or decreasing arrest numbers/crime.

I then break the overall analysis into three sub-stages, outlined below.

3.1 First stage - No Time Horizon

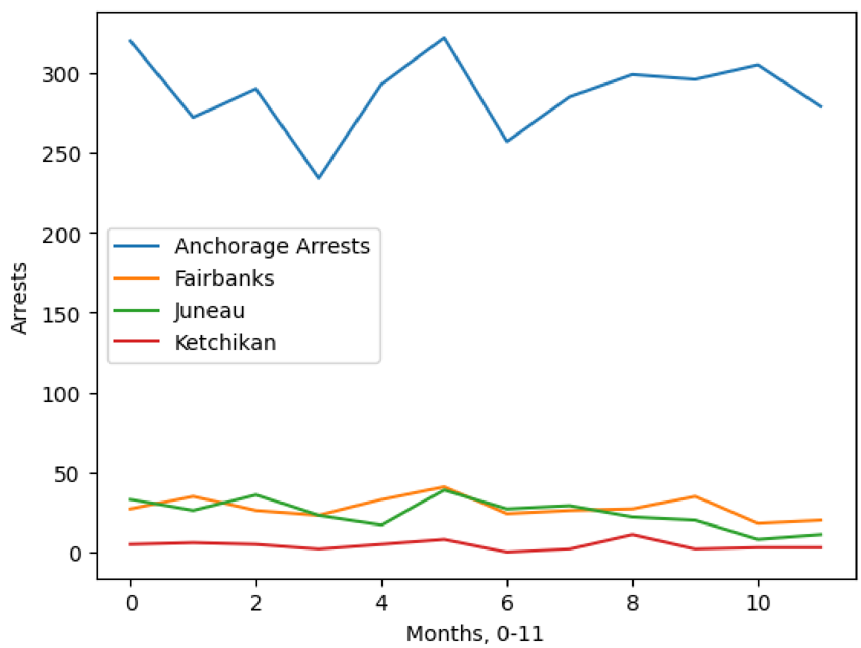
In the first stage, I investigate the effect of a police killing only on the arrests in that same month. Thus, the "post treatment" indicator column matches the "officer killed" indicator column since the treatment here is the death of an officer and each row corresponds to a month in a municipality. I then seek to estimate

$$\beta_0 + \beta_1 \text{shoot}_{it} + \beta_2 \text{Post}_{it} + \beta_3(\text{Treat}_{it} \times \text{Post}_{it}) + I + T + \epsilon_{it}$$

. However, notice that since the Post column and the "officer killed" column (treat column) are the same, the last two terms are actually the same, so the equation can be reduced to

$$Y_{it} = \beta_0 + \beta_1 \text{shoot}_{it} + \beta_2(\text{shoot}_{it} \times \text{killed}_{it}) + I + T + \epsilon_{it}.$$

Here, I and T are year and municipality fixed effects to account for intrinsic baseline differences in explanatory or response variables by municipality or year. To understand why we might care about these differences, consider the data on assault numbers in 2019 in the first four Alaskan municipalities.



Due to the size difference in these municipalities, the baseline levels of arrest are extremely different, even where the trends might be similar. To control for this I include place fixed effects I, as well as time fixed effects T.

The results of this first regression are shown below:

Here, the main coefficient of interest from our regression was $\beta_2 = 0.0289$, which isn't significant. However, the table as a whole suggests that even an officer's being *shot* in the line of duty has a significant impact on total arrests in the municipality (in particular, prompting a percent increase of 5.07 in arrest numbers), though additionally being killed may not make as much of an impact.

Table 2: Impact of LODD on arrests, same month

	Response Variable log(Total Arrests)
Shoot_{it}	0.0512^{***} (23.33)
$\text{Shoot}_{it} \times \text{Killed}_{it}$	0.0289 (0.83)
Constant	2.180^{***} (2447.66)
Observations	330456

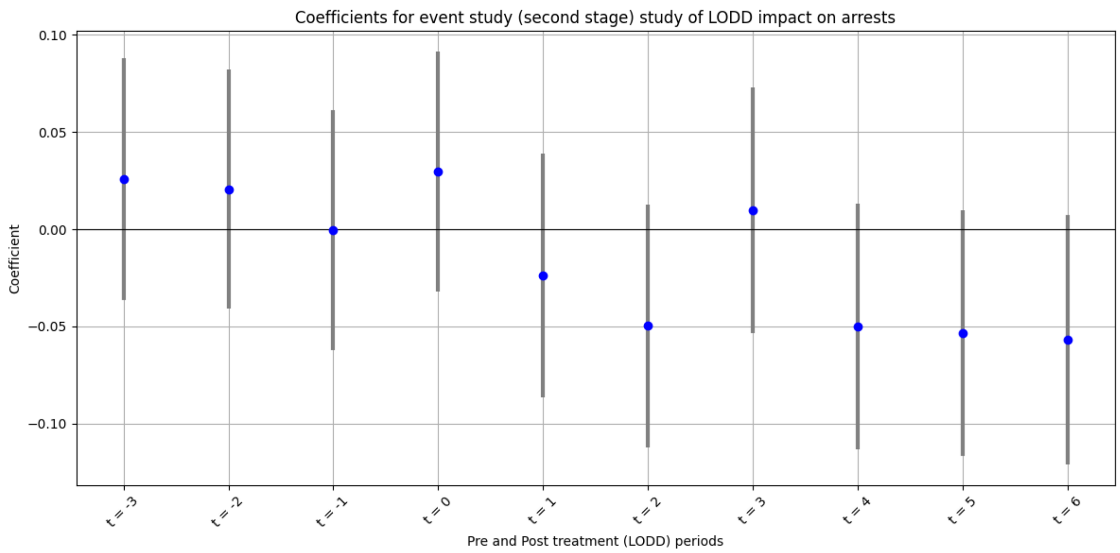
t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.2 Second stage - Event Study

In the second stage, I add post treatment indicators for 1, 2, ..., 6 months after a police killing; if there was a killing in month t in municipality i, the dummy Post_k takes on value 1 in month $t+k$ in municipality i; I also add pre-treatment indicators to confirm the absence of pre-trends or an anticipation-like effect. I only include a time horizon of six months because [3] found a significant effect only in the first few months. Since there are now post-treatment columns that are not directly equal to the killing column, the regression equation has post-treatment-inclusive interaction terms. Note that the interaction term coefficients represent the effects relative to untreated (shooting but no killing) baselines, and are thus the coefficients of interest. Taken together, the estimating equation becomes

$$\begin{aligned} Y_{it} = \beta_0 &+ \sum_{j=-3}^{6} \beta_{\text{Post, j}} \text{Post}_{j,it} + \sum_{j=-3}^{6} \beta_{PT,j} (\text{killed}_{it} \times \text{Post}_{j,it}) \\ &+ I + T + \epsilon_{it}. \end{aligned}$$

The results of this event study are shown below, with bars indicating confidence intervals for coefficient values; the complete table of values is given in the appendix.



Firstly, the graph confirms an absence of pre-trends, since there are no patterns in pre-treatment coefficients and none of them are significant; this affirms the validity of the parallel trends assumption, since there appears to be no pattern in the response variable prior to the introduction of the treatment. The apparent lack of pre-trends is also a positive sign for the robustness of the treatment effect estimation, as pre-trends may bias the coefficient estimates in the post-treatment months, though it is not a guarantee of unbiasedness. Additionally, it suggests that the treatment and control groups were similar with respect to other unobserved factors, reducing the risk of the estimated effect being affected by treatment-correlated unobserved variables.

Regarding the actual treatment effect however, I find no significant coefficients in the six post-treatment months, suggesting no treatment effect. The signs of the coefficients — ignoring $t = 3$, which seems like it might be an outlier, though the confidence interval encompasses both positive and negative values — are consistent with [3]: they are weakly negative, implying a decrease in arrests following the treatment and suggesting that risk-mitigating behavior is being practiced by the officers. The effect, while not significant in any of the months, seems more pronounced in months 4, 5, and 6, where the confidence interval almost fully falls below 0. It is possible that if I were to re-run the regression after adding more post-treatment time periods, I might find significance, since the general trend of the confidence intervals for the coefficients is a downward movement.

3.3 Third stage - Triple Difference

Finally, I add the size of the police force as a regressor, giving a triple difference estimating equation with the following interaction terms:

$$Y_{it} = \beta_0 + \beta_1 \text{Post}_{it} + \beta_2(\text{killed}_{it} \times \text{Post}_{it}) + \beta_3(\text{ForceSize}_{it} \times \text{Post}_{it}) + \beta_4(\text{killed}_{it} \times \text{ForceSize}_{it} \times \text{Post}_{it}) + I + T + \epsilon_{it}$$

The post term is now an overall term that simply indicates a treatment or not, using a window of 3 months pre and post treatment. The coefficients capture an average treatment effect.

I run this regression twice, with two differently calculations of force size; each represents a different interpretation that might affect the response differently.

The first is a measure of size relative to other police force sizes, which can alternatively be thought of as a measure of "absolute size." This can be interpreted as a measure of familiarity or fraternity within the police force. I predict that a smaller police force according to this metric will be tied to more arrests (negative coefficient) since the impact of a death might be felt more strongly in more absolutely small police forces.

The second metric is a measure of force size relative to municipality population. This can be interpreted as an indicator of how much cities are willing to invest in their police forces/the value of a police force to a city. Cities with a high ratio of force size to population might be more demonstrably supportive of the police or might offer other support resources to the police in the aftermath of police killings. I would expect that by this metric also, that a larger force size is tied to lower arrests in the post treatment period (negative coefficient). However, the magnitudes of the coefficients in the two cases might be different.

The results of the regressions are below:

Table 3: Impact of LODD and "absolute" force size on arrests

	Response Variable log(Total Arrests)
Post-Treatment$_{it}$ (PT$_{it}$)	-0.0608 (-1.29)
PT$_{it}$ $\times$ Killed$_{it}$	0.157* (2.40)
PT$_{it}$ $\times$ Force Size$_{it}$	0.00410 (0.08)
PT$_{it}$ $\times$ Force Size$_{it}$ $\times$ Killed$_{it}$	-0.100 (-1.41)
Constant	3.681*** (273.33)
Observations	1297

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Impact of LODD and population-relative force size on arrests

	Response Variable log(Total Arrests)
Post-Treatment$_{it}$ (PT$_{it}$)	-0.0640* (-2.36)
PT$_{it}$ × Killed$_{it}$	0.0914* (2.45)
PT$_{it}$ × Force Size$_{it}$	0.0131 (0.34)
PT$_{it}$ × Force Size$_{it}$ × Killed$_{it}$	-0.0368 (-0.70)
Constant	3.681*** (273.09)
Observations	1297

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We again find no significant effect due to either force size in the treated cases, or the treatment itself. Interestingly, when comparing the magnitudes of the coefficients in each notion of force size, it seems that there is a more significant reduction in arrest numbers when considering absolute force size, giving credence to the theory that familiarity with the officers may be the more important driver compared to resources in the police departments.

The lack of effect may be tied in part to the window of months used for indicating treatment ($t = -3$ to $t = 3$). Upon consideration of the event study's results, months $t = 1$ to $t = 3$ were the months with the least significant treatment effect coefficients; it is possible that if months $t = 4$ to $t = 6+$ were included, that some significance in the treatment effect coefficients may appear.

3.4 A comment about the null effects

With the exception of the coefficient for the effect of a shooting on arrests in the first-stage regression, all computed coefficients appear to be nulls, not significantly different from $\beta = 0$ H_os. The precision of this result is difficult to assess with certainty, but some contributing factors are the the imbalance of the dataset, large dataset size, and variability in the dataset.

The imbalance in the dataset — that is, the very small ratio of positive examples to total examples — makes it likely that the results are noisy nulls. In each year of data, there were around 40 - 50 officers killed, constituting a very small percent of the examples in the 10,000 municipalities and 12 months per municipality of data. This lack of positive examples could be resulting in high variability/standard errors in the estimates; it could also result in increased likelihood of a Type 2 error, underestimating the significance of the results. While in general, large dataset size increases precision, in this case, the large dataset size when taken together with the small number of positive examples may increase the noise in the result. The effect of this imbalance is present across all regressions and results.

4 Conclusions and Future Work

This paper contributes to work in policing and studying drivers of police behavioral response. I provide updated estimates with a sample restriction for the effect of a line of duty death on arrest numbers in the relevant municipality; I also introduce a heterogeneity, police force size, and study its effect when combined with the LODD treatment. I find no significant effects in either study, though the signs and magnitudes of the coefficients suggest that arrests go down and officers engage in risk-mitigating behavior; that this effect is *more* significant in later months and could possibly result in significant effects for months beyond the tested horizon; and that familiarity with the fallen officer is a bigger driver of downstream behavior change than, for instance, the resourcedness of the police department. While all results were found to be non-significant, the relatively small number of officers killed in the dataset contributes to a possibly noisy null, and may result in underestimations of the significance of the treatment effect.

In the future, I would first want to investigate the time trajectory beyond 6 months post-treatment (possibly 7 months to a year, or even simply a 6+ months indicator). Additional work on the possible effect of the arrests or policing behavior on the population could investigate potential reverse causality; it could also be worth investigating the types of arrest stops being made to lend credence to the theory of risk-mitigation suggested by the coefficients (eg. if stops are only happening for serious offences). It may also be worth studying demographic information like the race of the officer killed.

References

[1] Highest to lowest - prison population rate, 2023.

[2] Amelia Cheatham and Lindsay Maizland. How police compare in different democracies, 2022.

[3] Sungwoo Cho, Felipe Goncalves, and Emily Weisburst. The impact of fear on police behavior and public safety. *NBER Working Paper 31392*, 2023.

[4] Oeindrila Dube, Sandy Jo MacArthur, and Anuj K. Shah. A cognitive view of policing. *NBER Working Paper 31651*, 2023.

[5] Will Johnson. Community policing: Much more than walking a beat, 2018.

[6] Jacob Kaplan. Uniform crime reporting program data: Law enforcement officers killed and assaulted (leoka) 1960-2021, 2020.

[7] Jacob Kaplan. Uniform crime reporting (ucr) program data: Arrests by age, sex, and race, 1974-2021, 2020.

[8] Laurence Miller. Line-of-duty death: Psychological treatment of traumatic bereavement in law enforcement. *International Journal of Emergency Mental Health*, 9:13–23, 2007.

[9] Andreas Olden and Jarle Moen. The triple difference estimator.

Appendix - event study coefficients and SEs

Table 5: All coefficients and standard errors, event study

	Response Variable log(Total Arrests)
Post, -3$_{it}$ (interaction)	0.0259 (0.82)
Post, -2$_{it}$ (interaction)	0.0206 (0.66)
Post, -1$_{it}$ (interaction)	-0.000362 (-0.01)
killed$_{it}$ (interaction)	0.0296 (0.94)
post1$_{it}$ (interaction)	-0.0238 (-0.74)
post2$_{it}$ (interaction)	-0.0496 (-1.56)
post3$_{it}$ (interaction)	0.00966 (0.30)
post4$_{it}$ (interaction)	-0.0499 (-1.55)
post5$_{it}$ (interaction)	-0.0535 (-1.66)
post6$_{it}$ (interaction)	-0.0567 (-1.73)
Post, -3 (no interaction)	0.0306 (0.11)
Post, -2 (no interaction)	-0.192 (-0.53)
Post, -1 (no interaction)	-0.0494 (-0.17)
Post, 1 (no interaction)	-0.156 (-0.55)
Post, 2 (no interaction)	0.200 (0.53)
Post, 3 (no interaction)	-0.0777 (-0.24)
Post, 4 (no interaction)	-0.0213 (-0.06)
Post, 5 (no interaction)	0.195 (0.42)
Post, 6 (no interaction)	0.0321 (0.10)
Constant	2.192*** (2950.86)
Observations	330456

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The interaction terms (top coefficients) are the coefficients of interest and the ones reflected in the event study graphic, with a negative post value indicating a pre-treatment coefficient. The killed variable corresponds to $t = 0$, the month of treatment.