

# De-prosecution and death:

## A comment on the fatal flaws in [Hogan \(2022a\)](#)

JACOB KAPLAN\*

J.J. NADDEO<sup>†‡</sup>

TOM SCOTT<sup>§</sup>

September 20, 2022

### Abstract

In a manuscript recently accepted for publication in *Criminology & Public Policy*, [Hogan \(2022a\)](#) presents results from a synthetic control method analysis that suggests de-prosecution in Philadelphia in the mid to late 2010s resulted in a large increase in the number of homicides that occurred in the city. In this comment, we point out several critical errors in the analysis that when corrected flip the direction of the effect and render the author’s estimated effect null. Our primary concerns include the unjustified short pre-intervention period, a failure to correct for imbalance over covariates in the synthetic control models, the use of homicide counts instead of rates as an outcome, an inaccurate description of the data used, and an inadequate explanation of data cleaning procedures including missing data. We reproduce the author’s results after correcting for these issues and find no effect of de-prosecution on homicide. Thus, these flaws are fatal to the author’s findings and therefore the study should not be used to inform criminal justice policy. Considering the author’s unwillingness to share their data and code, we call for a greater dedication to open science and reproduction/replication in criminology.

**Note:** Updated on 9/20/2022 to include Appendix [C](#), which contains a reply to Mr. Hogan’s 8/23/2022 response to this article.

---

\*School of Public and International Affairs, Princeton University

<sup>†</sup>Massive Data Institute, McCourt School of Public Policy, Georgetown University

<sup>‡</sup>Institute for Technology Law & Policy, Georgetown University Law Center

<sup>§</sup>Division for Applied Justice Research, RTI International

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Issues with the methods</b>	<b>3</b>
2.1	Pre-intervention period . . . . .	4
2.2	Augmented synthetic control method . . . . .	6
2.3	Improper outcome specification . . . . .	9
<b>3</b>	<b>Issues with the data</b>	<b>13</b>
3.1	Source[s] of homicide data . . . . .	14
3.2	Incorrect homicide data and unit of analysis . . . . .	17
3.3	Incorrect outcome conceptualization . . . . .	18
3.4	Missing homicide data . . . . .	18
<b>4</b>	<b>Conclusion</b>	<b>19</b>
<b>A</b>	<b>Sample Selection</b>	<b>22</b>
<b>B</b>	<b>Mechanisms and competing theories</b>	<b>23</b>
<b>C</b>	<b>The Black Box of “De-prosecution and death” and the Importance of Reproducibility in Science</b>	<b>33</b>
C.1	Controlling for Clearance Rates . . . . .	34
C.2	Remaining Issues . . . . .	37
C.3	Conclusion . . . . .	42

# 1 Introduction

A recent paper published in the journal *Criminology & Public Policy* makes the argument that when a District Attorney’s Office (DAO) reduces the number of cases they prosecute murder will increase in that area. More specifically, when a DAO decreases prosecutions for all crimes—not just murder or similar crimes such as aggravated assault—murder, and only murder, will increase in response. This article, entitled “De-prosecution and death: A synthetic control analysis of the impact of de-prosecution on homicides” and written by Thomas Hogan—former federal prosecutor of the Eastern District of Pennsylvania and former District Attorney of Chester County, Pennsylvania—looks at Philadelphia as an example. Hogan found exactly as hypothesized, that when there are fewer prosecutions there are more homicides with an estimated “statistically significant increase of 74.79 homicides per year in Philadelphia during 2015-2019” [Hogan \(2022a\)](#).

We read this article with great interest as it studies an important and timely topic. During reading, we found many errors including misleading and inaccurate statements regarding the data and analytic methods, factual errors throughout the paper, and methodological problems. Most importantly, there is evidence that the author’s primary result—that de-prosecution increased homicide in Philadelphia relative to a synthetic control—is due to multiple fatal flaws in the analysis. For example, Hogan’s synthetic control model (SCM) failed to correct for a poor pre-trend fit, even though Alberto Abadie, one of the creators of SCM and an author of the R package that [Hogan \(2022a\)](#) uses and repeatedly cites as the authority on SCM, advocates to do so. We reran Hogan’s models applying this bias correction and not only do the results disappear, but they now point in the opposite direction. Corrected for bias, de-prosecution is related to a not statistically significant decrease in the number of homicides. Moreover, the author incorrectly used homicide counts instead of per capita rates as the outcome in their primary analyses. Despite claiming to control for population as an alternative, we reproduce [Hogan \(2022a\)](#)’s results without controlling for population as a predictor, suggesting the variable did not factor into the estimated effect.

We also show that if one restricts the donor pool used by the SCM to jurisdictions of similar populations to Philadelphia the results do not hold statistical—or substantive—significance. Similarly, if one expands the pre-treatment period the main result also does not hold.

Given these issues as well as others described in this comment, we believe that Hogan (2022a)’s findings are not reliable and should not be used to inform crime prevention policy. We present updated analyses that address the flaws in Hogan’s analyses and result in a different conclusion: there is no evidence that de-prosecution caused an increase in total homicides in Philadelphia. Because our primary concerns are with the methods and data used by the author, we emailed Mr. Hogan to ask him to share both the data and code used for this study. We emailed him on July 13, 2022 and again on July 23, 2022. He has yet to respond to our requests, so we use similar data and methods in our reproductions but were unable to exactly match the data given the inadequate level of detail provided in Hogan (2022a).<sup>1</sup> We strongly encourage authors to provide their data and code publicly following publication to allow for reproduction (Savolainen and VanEseltine, 2018). Our response to Hogan (2022a) covers two major areas of concern: 1) issues with the methods and 2) issues with the data.

## 2 Issues with the methods

The first set of issues we present involve flaws with the study’s methods, which we show result in an erroneous estimate of the impact of de-prosecution on homicide. Due to space consideration, we focus on the three most egregious flaws: 1) the short pre-intervention period, 2) the lack of bias correction for poor fit in the SCM analyses, and 3) the use of homicide counts instead of rates as the outcome. Not only are Hogan (2022a)’s results sensitive to correcting for these flaws, but when corrected, his estimated effect is rendered null. Because of this fact, we do not believe the study should inform policy debates on the consequences of various prosecution practices.

---

<sup>1</sup>The complete replication files for this response is available here: <https://doi.org/10.3886/E176021V1>

## 2.1 Pre-intervention period

Hogan (2022a) uses the SCM to estimate the relationship between de-prosecution and police-recorded city homicide counts. This estimator allows for estimation of treatment effects by creating a synthetic control unit comprised of a weighted average of control—or donor—units. The weights are selected to balance the treated unit and synthetic control in a pre-treatment time period. The current paper chooses a somewhat arbitrary pre-treatment time period of 2010-2014 (even though data exists for a much longer time period) and an arbitrary number of non-treated units (top 100 most populated municipal law enforcement jurisdictions out of roughly 12,700 municipal police departments) (United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics., 2017). Moreover, we do not believe that the analysis should be done at the law enforcement agency level, as the policy of interest occurs at the prosecutor level—which most closely aligns with the county (and may cover multiple agencies).<sup>2</sup> To illustrate how these selection decisions may impact the headline results we replicate the estimate of de-prosecution causing 74.79 additional homicides per year on average for 2015-2019. On our first pass, we use the Federal Bureau of Investigation’s (FBI) Uniform Crime Reporting Program’s (UCR) Supplementary Homicide Reports (SHR) data and apply the same difference-in-differences (DiD) methods described in Hogan (2022a).

Using SHR data for all donors and Philadelphia we find a statistically insignificant (at the conventional 95% level) difference of approximately 49 homicides per year between Philadelphia and synthetic Philadelphia (p-value = 0.055), representing a substantially smaller estimate than cited in Hogan (2022a) (see Figure 2). The discrepancy between our results and Hogan’s are likely because we use SHR total homicides for Philadelphia and all donor cities, which is what is stated in the text of Hogan (2022a). Inspection of Figure 2 in Hogan (2022a) and Table 1 in Section 3.1 of this article strongly suggests that Hogan (2022a) does not use SHR to measure the number of homicides in Philadelphia. Instead, it is more likely that he uses Philadelphia Police Department (PPD) data. Therefore, we re-

---

<sup>2</sup>A further discussion about the data implications of a county-level analysis can be found in Section 3.1.

place the SHR values with PPD values for Philadelphia and re-run the model.<sup>3</sup> As expected, we almost exactly replicate Figure 2 in [Hogan \(2022a\)](#) and our DiD estimate increases to 74.25. As mentioned above, Mr. Hogan did not provide his data and code (e.g. model specifications), so our results differ slightly but a statistically indistinguishable amount. We produce an estimate of 74.25 compared to 74.79 in [Hogan \(2022a\)](#) with almost identical donor weights: Detroit (46.8% compared to 46.8% in [Hogan \(2022a\)](#)), New Orleans (33.6% compared to 33.4% in [Hogan \(2022a\)](#)), and New York City (19.6% compared to 19.8% in [Hogan \(2022a\)](#)).

After reproducing [Hogan \(2022a\)](#)’s main result, we set out to address several major methodological concerns. Because we are able to nearly perfectly replicate the main results in [Hogan \(2022a\)](#) using the same donor cities with almost identical weights, the results we present in the following sections are due to differences in what outcome variation is used (i.e., raw versus residuals accounting for differences in predictors) and not differences in comparison cities between our analysis and [Hogan \(2022a\)](#). Said another way, our results still use Detroit, New Orleans, and New York City, but correct for differences in population, homicides cleared, and homicide clearance rates, while the results in [Hogan \(2022a\)](#) only looks at raw differences in total homicides. We note again that this exercise in recreating his data and results could have been avoided if [Hogan \(2022a\)](#) had made his code and data publicly available or had provided them to us when we asked.

First, we found [Hogan \(2022a\)](#)’s use of 2010 as the start of the pre-intervention period surprising given that the SHR data go back to the late 1970s and statisticians have clearly noted that a short pre-intervention period can result in a biased SCM estimate ([Abadie \(2021\)](#), p. 413; [Abadie et al. \(2010\)](#)). Importantly, [Hogan \(2022a\)](#) does not explain his

---

<sup>3</sup>Importantly, the PPD measures their homicides in victims not incidents. There can be more than one victim in a homicide incident so these numbers are not equivalent. After running one analysis using total number of homicide victims and one using total number of homicide incidents, it is likely that [Hogan \(2022a\)](#) used SHR incidents for the donor cities. When using incidents we recreate his chosen donor cities with near identical donor weights; using victims we find different donor cities. While we believe that incidents are likely the correct unit of observation based on [Hogan \(2022a\)](#)’s logic connecting de-prosecution to homicide, because the PPD data that [Hogan \(2022a\)](#) uses measures total victims, Philadelphia’s total homicides are inflated relative to potential donors in the dataset.

choice in date or how it relates to the equivalency between the treated and synthetic control units.<sup>4</sup> To test the sensitivity of Hogan (2022a)’s estimate to different pre-intervention start dates, we systematically vary the start date of the pre-intervention period from 2000 to 2010. We allow the SCM to choose different weights for each new pre-intervention period sample and present the resulting DiD estimates in Figure 1. Three points to take from this analysis include 1) the estimated effect is sensitive to the duration of the pre-intervention period, 2) the selection of 2010 maximizes the DiD estimate, and 3) the synthetic control unit’s composition differs meaningfully across dates.<sup>5</sup>

## 2.2 Augmented synthetic control method

Our next major concern involves the arguably poor pre-intervention fit in the author’s main SCM analysis and the fact that nothing was done to correct for this bias. Specifically, in spite of a generous y-axis, upon inspection of the pre-trend fit of Figure 2 of Hogan (2022a), one notices a clear pre-trend divergence in the period directly before treatment. In fact, the divergence is in the direction of positively biasing the final results as real Philadelphia is trending upward while the synthetic control’s trend is flat. We believe that this divergence warrants at least testing for whether the results are robust to bias-correction methods. Notably, the author cites one of the creators of the SCM—Alberto Abadie—multiple times but fails to check if the results are robust to a recent refinement to the estimator in Abadie and L’Hour (2021).<sup>6</sup> This methodological refinement allows the researcher to correct for poor

---

<sup>4</sup>There is also confusion by Hogan about what units his data are in. When discussing his primary result, the 2015 start of the post-period, he says in footnote 10 that “With the pre-period match coded to end in 2014, the algorithm actually allows the real-life divergence to begin in mid-2014.” This is not possible, however, as the data he uses throughout the paper are annual, not monthly, homicide counts. No algorithm can measure mid-year data when the unit of analysis is yearly. It is clear throughout the paper that analyses were based on annual counts rather than monthly counts. For example, in every figure and table in the paper the data are shown as annual numbers. In every regression result Hogan describes the effect size as homicides per year, never as homicides per month or quarter as would be appropriate when using sub-year units.

<sup>5</sup>Depending on the start date used, the synthetic unit is composed of different combinations of 9 unique cities, ranging from 3 to 5 donors.

<sup>6</sup>We call attention to Abadie and L’Hour (2021), but recognize that this method was also pioneered by Ben-Michael et al. (2021).

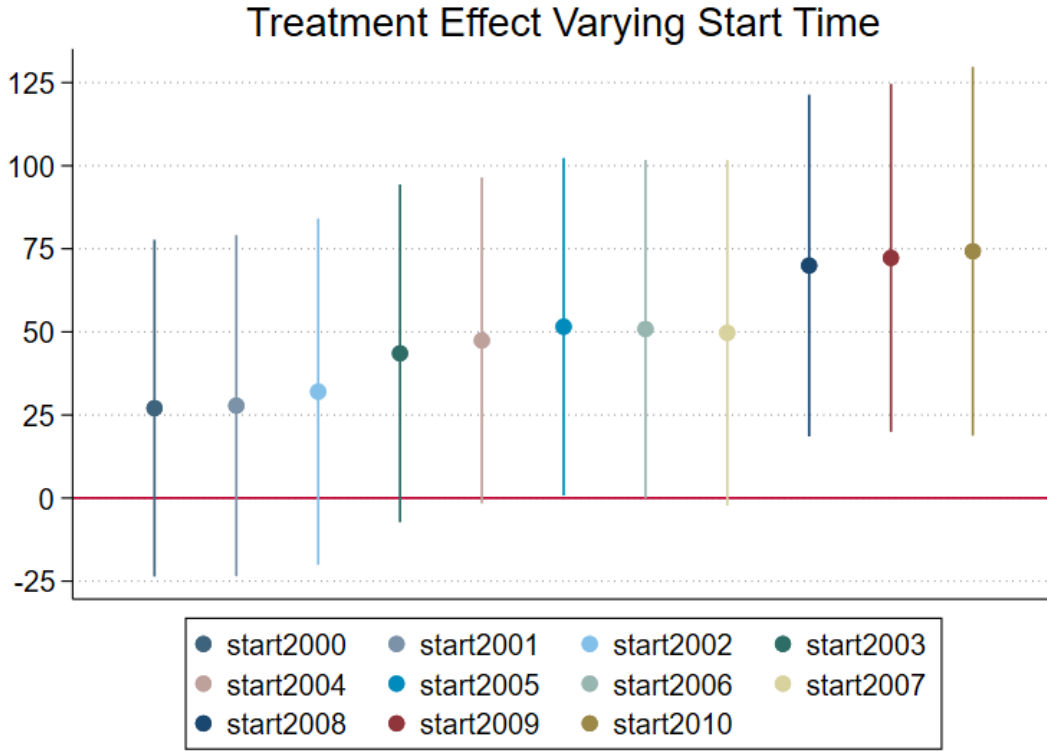


Figure 1: DiD estimates using same data and SCM method in [Hogan \(2022a\)](#) but varying the start date of the pre-treatment period. We used PPD data accessed [here](#) to correct Philadelphia’s homicide count for 2007-2019 (all years available) as in [Hogan \(2022a\)](#), and SHR from 2000-2006. Confidence intervals are at the 95% level.

pre-trend fit—and in fact is mentioned in [Abadie \(2021\)](#) (p. 418), which is cited multiple times in [Hogan \(2022a\)](#). To illustrate how these bias correction methods may influence the main results we compiled a similar dataset and replicated [Hogan \(2022a\)](#)’s Figure 2 in Figure 2 below. Note again, we find a DiD estimate of 74.25 additional homicides per year, statistically indistinguishable from the 74.79 estimate in [Hogan \(2022a\)](#).

Using the same data, we apply the bias-corrected SCM developed by [Abadie and L’Hour \(2021\)](#); [Ben-Michael et al. \(2021\)](#) using the *allsynth* package in Stata.<sup>7</sup> Figure 3 plots the difference in real and synthetic Philadelphia for the traditional SCM results (same as Figure 2) and bias corrected results. As is clear from the figure, using the bias-correction method

<sup>7</sup>See [Wiltshire \(2022\)](#) for details on the software. To be specific, we use the default settings which use OLS regression to estimate the bias due to imperfect pre-period fit.



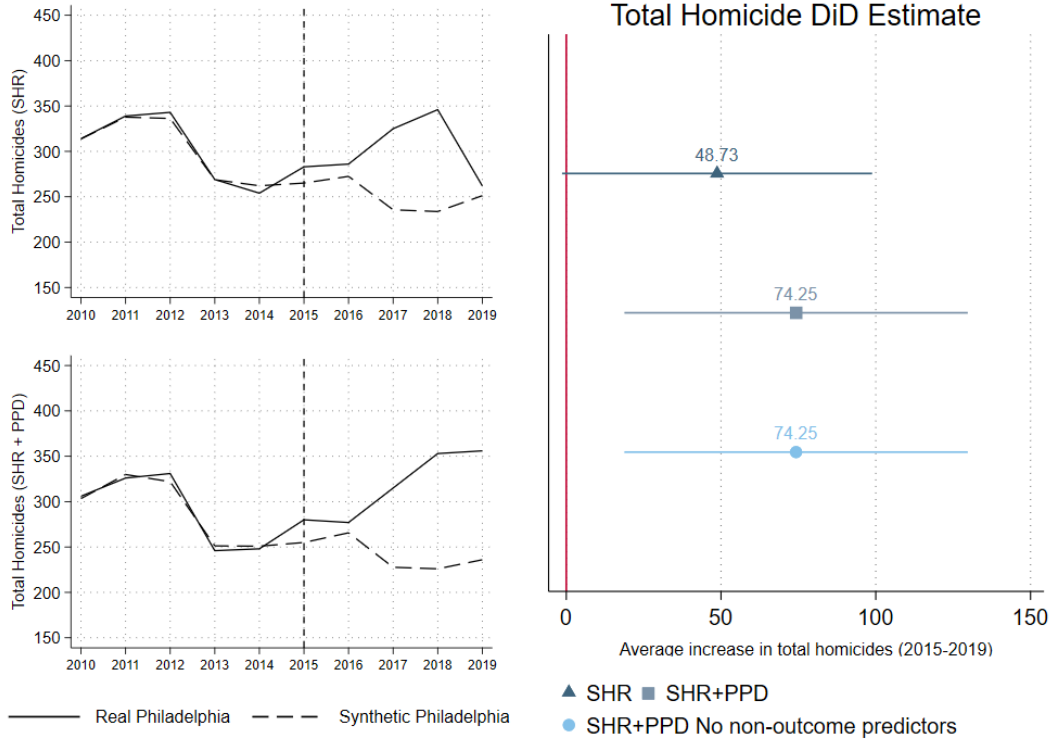


Figure 2: Replication of Figure 2 in Hogan (2022a): (top left) using only SHR data; (bottom left) using SHR data for all donors and PPD data for Philadelphia; (right) Coefficients from DiD estimation using SHR and SHR + PPD data with same model as in Hogan (2022a) (except that we did not use median income as a predictor) and SHR + PPD with no predictors except the pre-treatment outcomes. Confidence intervals are at the 95% level.

flips the sign of the estimated treatment effect. Specifically, the DiD estimator using bias-corrected SCM implies that the average effect on homicides between 2015-19 is -46.22 (p-value = 0.128), compared to the estimate of 74.25 (p-value = 0.015) produced by the traditional SCM that is not corrected for bias in pre-intervention fit (see right panel in Figure 3). However, it should be noted that the effect of bias correction depends crucially on the functional form of the predictors. As Hogan (2022a) does not clearly indicate the exact specification he is using but only provides the variables that are used, we allow these variables (sans median income) to enter as flexibly as possible by including the values for 2010, 2011, 2012, 2013, and 2014.<sup>8</sup>

<sup>8</sup>The patterns we find are not substantively different if we use the other extreme and only include the average predictor values over the pre-treatment period. This would assume that predictor effects are constant

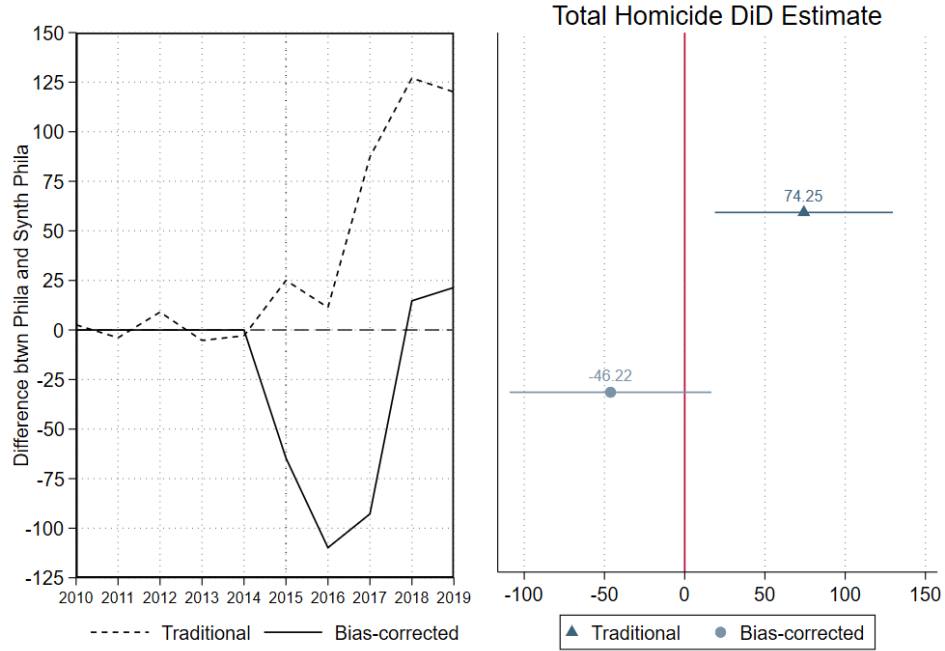


Figure 3: (Left) Gap in total homicides between real and synthetic Philadelphia using population, number of cleared homicides, and the homicide clearance rate as predictors using traditional SCM and bias-corrected SCM. (Right) DiD estimate of average change in homicide using both methods.

In summary, Figure 2 illustrates a concerning difference in pre-treatment trends in the synthetic control and Philadelphia. As suggested in Abadie (2021), in the case of poor pre-intervention fit between the treated and synthetic control units, one should apply a bias-correction to the SCM like the one developed by (Abadie and L’Hour, 2021; Ben-Michael et al., 2021). When we apply these methods to the same data, the estimates flip signs (Figure 3), a result that calls into question the main results of the paper.

## 2.3 Improper outcome specification

Almost any level measurement of an area-wide outcome will be highly correlated with the population living in that area. Therefore, it is common practice to use per capita rates to account for differences in population size when testing hypotheses. In the current study, the

---

over time.

research question centers around if de-prosecution causally affects violent crime (specifically homicide). At a bare minimum, one would expect to control for population size when attempting to answer this question using between city variation.<sup>9</sup> Hogan (2022a) instead uses raw total homicide counts as the outcome of interest and fails to account for variation in city population size. Hogan (2022a) makes three arguments for his choice in outcome. First is that including population as a balancing predictor variable in the SCM alleviates the need to account for city population differences using per capita rates. However, we illustrate in this section that this is not the case. Second, the author claims that population size is not relevant because the sample is limited to large cities. Yet, this is unsatisfactory given that during the observation window city population size ranges from less than 300,000 to over 8,500,000 residents. Third, the author vaguely cites an argument made by Abadie (2021) for restricting the donor pool to comparable units out of context as suggesting that if the treated unit differs greatly from the donor pool on one outcome, it is good practice to ignore that outcome and use a less suitable outcome that is more comparable across units instead. Based on Abadie (2021), we believe Abadie would argue this action is likely to result in interpolation bias, which should be addressed by “restrict[ing] the donor pool to units with characteristics that are similar to the affected unit” (p. 409; see also Abadie et al. (2010)). Frankly, if there is no weighted combination of cities that is similar to Philadelphia in terms of the homicide rate, then one should not use the SCM to estimate a policy effect on homicide for Philadelphia. Below, we present an empirical basis for our concern over Hogan (2022a)’s choice of outcome.

Figure 2 above shows the results using what we believe to be the same predictors (the author does not explicitly state the actual model used anywhere in the text). We then remove all predictors, including, notably, city population size, excepting the pre-intervention period homicide counts for each year and find that the results remain identical (i.e. same donor weights and identical point estimates/standard errors around the DiD estimate, see

---

<sup>9</sup>For example, one could residualize out population from homicide counts before implementation of SCM.

last coefficient right panel of Figure 2). To reiterate, the homicide count based estimates in Hogan (2022a) are not factoring in city differences in resident population size despite it being included as a covariate. While at first surprising, upon closer inspection it becomes easy to understand the irrelevance of the non-outcome predictor variables. The SCM uses predictor variables to balance pre-treatment outcomes.<sup>10</sup> Therefore, it is foreseeable that the predictors that receive all the weight in the balancing of the pre-intervention period outcomes are the annual outcomes.<sup>11</sup> Said another way, when attempting to find donors that most resemble Philadelphia with respect to total homicides, total homicides is the best (only) predictor that receives weight. As shown in Section 2.2 when we utilize bias-corrected SCM—which corrects for imbalance in predictor variables (e.g., population)—the main estimates flip sign. Because the count-based analyses in Hogan (2022a) do not account for large differences in city population size, which is strongly associated with levels of crime and violence, we discount them and use the more valid measure of homicide rates in further analyses.

Importantly, although Hogan (2022a) also provides an estimate using homicide rates instead of counts (pp. 19-20), they do not test this for the main specification in which treatment begins in 2015 nor do they subject the estimate to robustness checks. Instead, they choose to use a single treatment period of 2017. Following Hogan (2022a), we employ SCM on the homicide rate using the same predictors as the original model and using the treatment period of 2017. As noted in Hogan (2022a), this results in almost every potential donor receiving non-zero weights. As noted in Abadie (2021), a lack of sparsity (i.e. non-zero weight on more than a handful of donors) is indicative of a failing of the SCM with the solution likely being non-unique. Therefore, it is recommended that the donor pool be restricted and/or more predictors be used to find a sparse set of donors. Hogan (2022a)

---

<sup>10</sup>We assume that the model used by Hogan (2022a) uses homicide totals, population, total homicides cleared, and homicide clearance rate in 2010, 2011, 2012, 2013, and 2014 (i.e. all pre-treatment periods) and the median income over 2010-2014 to select donor cities that most resemble the homicide totals in Philadelphia. We are forced to assume as the model is never explicitly defined in Hogan (2022a).

<sup>11</sup>In fact, upon inspection of the variable weighting matrix produced by the SCM, we find that all non-outcome variables received weights ranging from 1.125e-12 to 4.01e-9. In contrast, the annual total homicide counts received weights of 0.237, 0.240, 0.217, 0.150, and 0.157 for 2010-2019, respectively.

instead notes the failing of the SCM but proceeds to interpret the results, something we caution against. Regardless, we present the bottom left panel of Figure 4 which replicates Figure 7 in Hogan (2022a) but caution that the results should **not** be interpreted. We again caution that there is a pre-trend that is salient.

What is far more troubling is that the author does not check the results using the preferred/main treatment period of 2015. As shown in Figure 4, we replicate the rate-based analysis using both the author-preferred date of 2015 in the top-left panel and the original date of 2017 in the bottom-left panel.<sup>12</sup> Turning to the DiD estimates provides a potential reason for the author’s choice of 2017 as the intervention year for this analysis: using 2015 (or 2016) as the treatment period cuts the estimate by over 50% (nearly 50%) compared to 2017 and the estimated effect is not statistically significant (p-value = 0.163 for 2015, p-value = 0.128 for 2016). However this is likely due to the same reason why the estimate should not be interpreted, the SCM places weight on every donor. Thus, it is sensitive to slight changes to the donor pool and/or choice of pre-treatment period (i.e., it is unstable). Regardless, the more important result from this exercise is that as the main specification—that produces the headline results—for which Hogan should be testing “robustness” is using 2015 as the treatment period we find it concerning that he decides to use 2017 to test the homicide rate when 2015 produces null results. While we stress that any results produced by these models should not be interpreted, we further illustrate the dubious nature of the findings by applying the same bias-correction methods used in Section 2.2. Figure 5 presents these results and shows—as was the case with homicide counts—that accounting for imbalanced covariates removes a positive bias, causing a reversal of the findings in Hogan (2022a).

In summary, we believe further work should be done to find a refined model using homicide rate as an outcome that produces sparse donors and strong pre-treatment fit.

---

<sup>12</sup>For 2017, we find a somewhat larger estimate of 5.37 homicides per 100,000 population (p-value = 0.002) than the 4.06 estimate in Hogan (2022a). Because Mr. Hogan did not respond about sharing his data and code, and because of the lack of documentation in Hogan (2022a) regarding the study’s methods, there are slight differences in our sample that led to minor differences across manuscripts. We note that even with our larger effect size in 2017, the estimate still becomes null when 2015 (or 2016) is used as the intervention date.

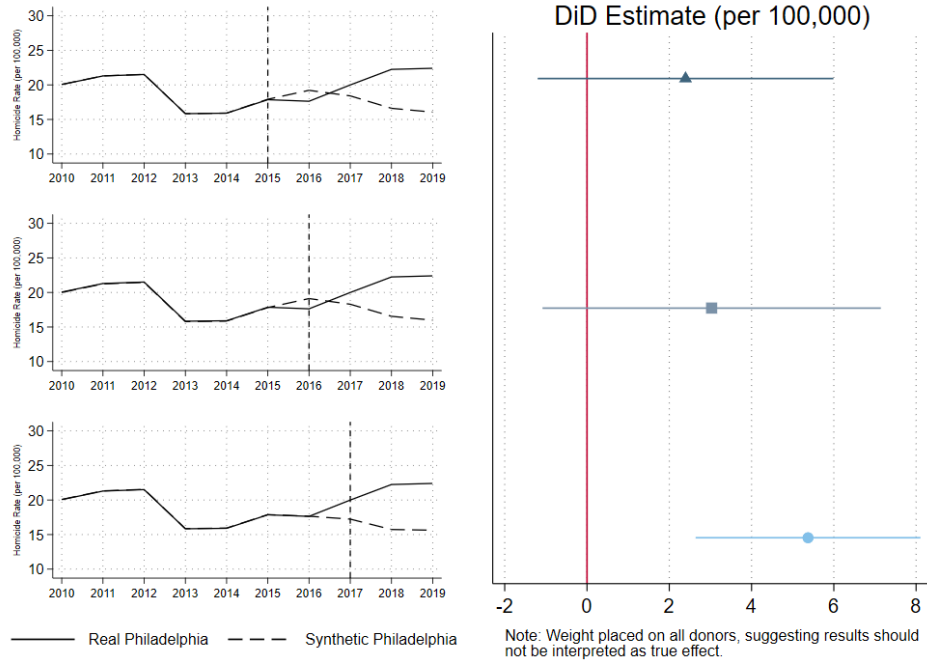


Figure 4: Left: Replication of Figure 7 in Hogan (2022a), varying the treatment time from 2015 (top), 2016 (middle), and 2017 (bottom). Right: DiD estimate of average increase in homicide rate due to treatment.

Moreover, we find it concerning that the author decided to use 2017 instead of 2015 (the specification used to produce the headline result) when checking robustness to changing the outcome to a more intuitive outcome—the homicide rate. Especially because when using the author’s preferred policy date of 2015, we find statistically insignificant results compared to the statistically significant (and much larger in magnitude) estimates using 2017 as the intervention year.

### 3 Issues with the data

In addition to our concerns with the outcome and estimator used in Hogan (2022a), we identified multiple problems with how the author described the data including their source and how missing data were addressed. Although we view these issues as being less critical to the study’s findings compared to issues raised in the prior section, they raise legitimate

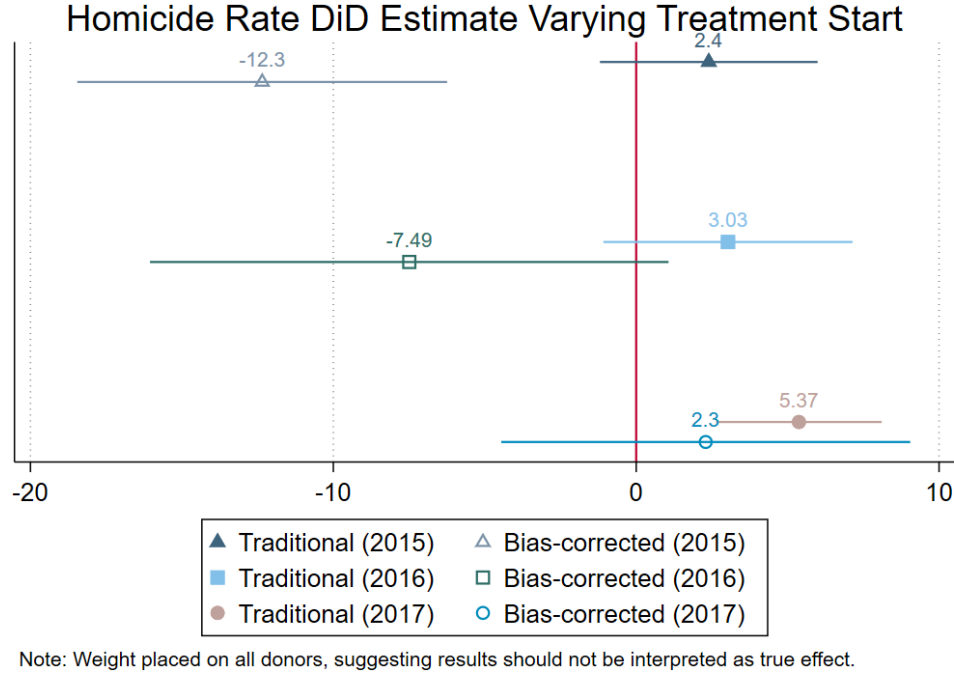


Figure 5: DiD estimate of average increase in homicide rate due to treatment, varying the treatment period from 2015 (top) to 2017 (bottom). Solid (hollow) markers represent results from traditional (bias-corrected) SCM.

concerns regarding the quality of the author’s work and the accuracy of statements in the manuscript. There are four primary issues with the homicide outcome: 1) a misrepresentation of the multiple data sources used for different city-years, 2) a use of incorrect data and unit of analysis, 3) an incorrect conceptualization, and 4) an inadequate description of missing data procedures.

### 3.1 Source[s] of homicide data

In the data section, [Hogan \(2022a\)](#) states that he uses the FBI’s SHR data to measure his dependent variable, the number of homicides that city police departments record each year from 2010 through 2019. He goes on to cite the source of the data: (Kaplan, J. (2021). Uniform Crime Reporting Program data: Supplementary Homicide Reports. Inter-university Consortium for Political and Social Research) which is a beneficial though uncommon prac-

tice in much social science research. Though [Hogan \(2022a\)](#) explicitly states that he uses SHR data to measure homicide, it is actually quite unclear what data he uses across his analyses. For example, there is evidence that [Hogan \(2022a\)](#) used three separate data sources to measure homicide in his manuscript: the SHR, the FBI’s Offenses Known and Clearances by Arrests data (which we describe in detail below), and local agency data.

For his primary analyses, it appears that Hogan used the SHR to measure homicide counts for the donor cities and data from the PPD’s [web page](#), which lists the number of homicides each year from 2007 onward, to measure homicide counts for Philadelphia. In Table 1 below, we show the number of homicides in Philadelphia from four sources: [Hogan \(2022a\)](#), the PPD website, the Offenses Known data, and the SHR data that we divide into all victims and only murder victims.<sup>13</sup> Given that PPD reported an incomplete annual homicide count to the FBI in 2019, the use of the SHR would have been problematic. However, it may also be inaccurate to compare PPD’s self-reported data to data from the SHR if homicide is measured differently across data sources. If the author believes that local city data is superior to SHR then that decision should be stated clearly and should apply for all cities with available data, not only for Philadelphia. The author did not discuss this potential and in fact did not explicitly state that Philadelphia’s homicide data did not come from the SHR as stated in the methods section.

At other times in the manuscript, [Hogan \(2022a\)](#) does not use SHR data and instead relies on the Offenses Known data to measure homicide. For instance, the solid black line in Figure 10 (shown in Figure 6) shows Baltimore’s homicide counts compare to a synthetic control of nonprogressive cities.<sup>14</sup> To focus on only two years of interest, 2014 (the year in which the post-period starts) shows homicides at a little over 200 and by 2015 that number spikes to a little below 350. As explained in the data section, these numbers are

---

<sup>13</sup>Importantly, police-recorded crime data can be measured based on the number of victims or incidents. [Hogan \(2022a\)](#) does not state his unit of observation and there is evidence it is inconsistent between Philadelphia (# of homicide victims) and the donor cities (# of homicide incidents). We present victim-based numbers here following the PPD’s unit of observation.

<sup>14</sup>Figures 9 and 10 are meant to show different information (Chicago and Baltimore, respectively), but the figures are identical (Baltimore). It seems like the author mistakenly used the same figure in both instances.



Table 1: Total Homicide Count in Philadelphia from Multiple Sources

Year	Hogan (2022a)	Philadelphia Police Website	FBI Offenses Known Data	FBI SHR data (all homicides)	FBI SHR data (all murders)
2010	306	306	306	321	306
2011	326	326	324	347	324
2012	331	331	331	354	331
2013	246	246	254	276	247
2014	248	248	248	258	253
2015	280	280	281	287	280
2016	277	277	274	289	273
2017	315	315	317	325	316
2018	353	353	351	357	351
2019	356	356	266	270	265

ostensibly based on SHR data. However, they are substantially far from SHR numbers where the Baltimore [city] Police Department reported 193 and 282 homicides in 2014 and 2015, respectively. When using the Offenses Known data, these numbers are 211 in 2014 and 344 in 2015, the same as shown in the graph.

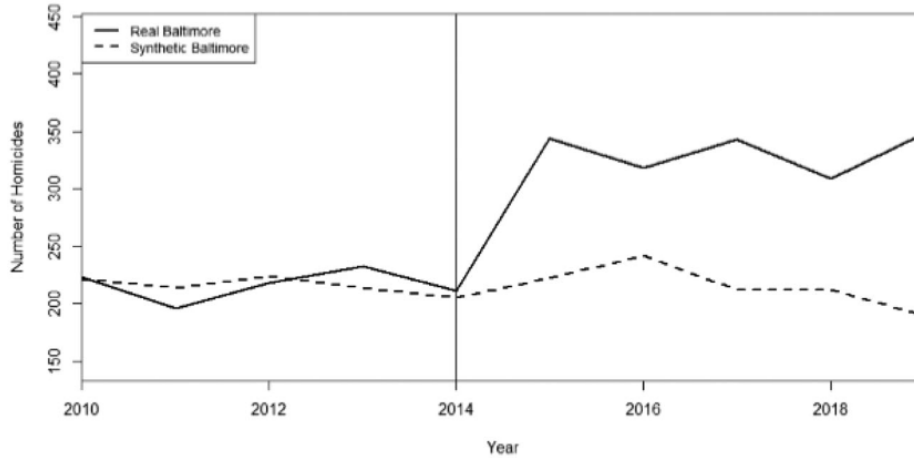


Figure 6: Figure 10 from Hogan (2022a)

Finally, in addition to using homicide counts from PPD’s website, Hogan (2022a) states that he obtains homicide counts that were missing from the SHR from local agency websites. Importantly, he fails to 1) describe how frequently this occurred, 2) to cite the websites, and 3) to discuss any steps taken to ensure the equivalence in homicide measurement across data sources that may not be alike. In sum, despite a statement in the methods section that the

homicide outcome was sourced from the SHR, we identified three distinct data sources used to measure homicide in the author’s main analyses.

### 3.2 Incorrect homicide data and unit of analysis

Even if [Hogan \(2022a\)](#) had used SHR data throughout the paper, it is unclear why he would use this data collection rather than the Offenses Known and Clearances by Arrests (also called the “Return A” or “Summary Reporting System”) data collection, which, like SHR, is part of the FBI’s UCR Program data [Kaplan \(2022\)](#). Importantly, the Offenses Known dataset is a more complete measure of police-recorded homicide; SHR data is often an undercount of homicides compared to Offenses Known, even for the same agency reporting data to both data collections ([Kaplan \(2022\)](#), Chapter 6 Supplementary Homicide Reports, Figure 6.1).

In fact, it is even more unclear why [Hogan \(2022a\)](#) did not use county homicide counts and rates from the Centers for Disease Control and Prevention’s (CDC) Multiple Causes of Death data given that it is more complete than police-recorded homicide counts in the SHR and Offenses Known data ([Kaplan \(2022\)](#), Chapter 6 Supplementary Homicide Reports, Figure 6.1) and the county is a more valid unit of observation given that DAOs prosecute cases within counties not cities. For example, [Hogan \(2022a\)](#) classifies “New York City” as “Middle”, however the city contains five separate counties with five different DAOs. Additionally, although multiple law enforcement agencies may exist within a DAO’s jurisdiction, [Hogan \(2022a\)](#) does not disclose how—or even if—he collapses law enforcement agencies into DAO jurisdictions. Researchers must exercise great caution when collapsing law enforcement agency level data to the county level ([Maltz and Targonski, 2002](#); [Pridemore, 2005](#)). Based on our reading, [Hogan \(2022a\)](#) conducted all analyses using single city police department data as opposed to county level death certificate data, which we view as the wrong data and unit of analysis.

### 3.3 Incorrect outcome conceptualization

Both the SHR and Offenses Known data provide monthly counts of the number of police-recorded 1) murders and non-negligent manslaughters, 2) negligent manslaughters, and 3) justifiable homicides (U.S. Department of Justice Federal Bureau of Investigation, 2004).<sup>15</sup> These offense distinctions are valuable given that the three homicide types likely have unique etiologies. Based on the lack of detail in the Data Used section in Hogan (2022a), it appears these three unique types of homicide were combined to form the author’s outcome. Typically, theory sections are used to describe causal models and methodological decisions like the conceptualization of the outcome. Because this section is absent in Hogan (2022a), we do not know the author’s thinking for combining these distinct homicide types. Our thinking is that any causal pathways linking de-prosecution to an increase in killings due to negligence or justifiable circumstances would be convoluted and a more appropriate outcome would be murder, or even interpersonal violence more broadly. At a minimum, Hogan (2022a) should have explained that his homicide outcome combined these three distinct offense types. Preferably, he would justify the conceptualization.<sup>16</sup>

### 3.4 Missing homicide data

Another problem with the manuscript is the author’s description of the methods used to address missing homicide counts. For example, in footnote 4 Hogan (2022a) says that “Where any homicide data for a specific year and city were not listed in the UCR/SHR, the information was retrieved from publicly available sources for specific police jurisdictions.” Importantly, Hogan (2022a) does not say how often this occurred, for which cities and years it occurred, nor does he cite the websites in which data were collected to replace missing val-

---

<sup>15</sup>Although the Offenses Known dataset does not report justifiable homicides in the “actual” number of crime offenses, it can be determined by summing the “actual” number of homicides with the “unfounded” number. For more detail on Offenses Known reporting please see (Law Enforcement Support Section and Crime Statistics Management Unit, 2013).

<sup>16</sup>We note that the inclusion of negligent manslaughter and justifiable homicide with murder and non-negligent manslaughter is unlikely to substantively alter findings as both are rare relative to murder Kaplan (2022), Chapter 6 Supplementary Homicide Reports (SHR), Table 6.3).

ues. Not only does this hinder efforts at reproduction, but it makes it impossible to quickly judge the quality of the analysis and associated results.

Moreover, although Hogan vaguely explains how he addressed missing city-years of homicide data, he does not discuss how he addressed missing months of data among agencies that reported at least one month of data. For example, PPD reported approximately 90 more homicides on their website than they did in the SHR in 2019. This is due to missing months of data for that agency in the SHR. Although we previously explained that the author used PPD data instead of SHR data in his analyses, there was no discussion of whether he checked for missing months of data for donor cities across the observation period. Although in a robustness check [Hogan \(2022a\)](#) states “approximately 20% of the data that I used to calculate burglary and robbery offenses had to be imputed from other years”, he does not explain whether those missing data included missing months or missing years of data, nor does he describe his method of imputation. The inadequate description of his methods for addressing missing data compound his inaccurate description of the data used to measure homicide, which results in the unpleasant situation of not knowing which statements to trust and which analyses to put faith in when reading the manuscript.

## 4 Conclusion

[Hogan \(2022a\)](#) ends by saying that “Every criminal justice policy—from ban-the-box to stop-and-frisk—should be evaluated for both intended and unintended downstream effects.” We strongly agree. All policies should be evaluated thoroughly. This also applies to all research papers, and we believe that we have done a thorough evaluation of this one. We find that there are numerous methodological, data, and factual errors in this paper which lead us to believe that its findings cannot be trusted. In this comment, we present what we view as the most critical flaws in the study, which we feel requires one to discount [Hogan \(2022a\)](#)’s presented analyses. These include an unjustified short pre-intervention period, a

lack of bias correction in the SCM analyses, the use of homicide counts instead of rates as the outcome, a misrepresentation of the data, and an inadequate description of the methods applied to clean the data including addressing missing values.

Because addressing the major methodological issues in [Hogan \(2022a\)](#) rendered his once positive and statistically significant estimated effects null, we describe these errors as fatal flaws. We identified multiple other concerns with the study that we view as more minor but nonetheless add to the skepticism we share over the credibility of the manuscript. We addressed some of these concerns to the best of our ability not having the author’s data and code in updated analyses presented here and found no evidence to support the author’s claim that de-prosecution resulted in a future increase in homicide in Philadelphia. That is not to say that there is no relationship, just that the purported relationship in [Hogan \(2022a\)](#) does not hold up when correcting for major flaws in the original analyses.

This paper covers an important topic, and this topic requires diligent research. We hope that other researchers continue to study the relationship between prosecution and crime generally and progressive prosecutors and homicide specifically. While doing so, we encourage criminologists to preregister their studies, especially if there is an appearance of a conflict of interest, and to publish their code and data upon publication of a scientific manuscript to promote reproduction in support of evidence-based crime policy.<sup>17</sup> We also encourage the American Society of Criminology and other criminology journal administrators to adopt policies that require publishing authors to make their data and code publicly available, at least for the purpose of reproduction.

---

<sup>17</sup>This is especially true if one’s data are publicly available and simply involve slight changes to the data such as imputation of missing values. Maybe even more so if one of the authors requesting the data generated it for public use.

## References

- Abadie, Alberto, “Using synthetic controls: Feasibility, data requirements, and methodological aspects,” *Journal of Economic Literature*, 2021, 59 (2), 391–425.
- , Alexis Diamond, Hainmueller, and Jens, “Synthetic control methods for comparative case studies: Estimating the effect of California’s Tobacco control program,” *Journal of the American Statistical Association*, 2010, 105 (490), 493–505.
- and Jérémy L’Hour, “A Penalized Synthetic Control Estimator for Disaggregated Data,” *Journal of the American Statistical Association*, 2021, 116 (536), 1817–1834.
- Agan, Amanda Y., Jennifer L. Doleac, and Anna Harvey, “Misdemeanor Prosecution,” 2021.
- Avdija, Avdi S, Christian Gallagher, and DeVere D Woods, “Homicide Clearance Rates in the United States, 1976-2017: Examining Homicide Clearance Rates Relative to the Situational Circumstances in Which They Occur.,” *Violence and victims*, 2 2022, 37 (1), 101–115.
- Ben-Michael, Eli, Avi Feller, and Jesse Rothstein, “The Augmented Synthetic Control Method,” *Journal of the American Statistical Association*, 2021, 116 (536), 1789–1803.
- Hogan, Thomas P., “De-prosecution and death: A synthetic control analysis of the impact of de-prosecution on homicides,” *Criminology & Public Policy*, 2022, pp. 1–46.
- Hogan, Tom, “DE-PROSECUTION AND DEATH: A Cordial Reply to Kaplan, Naddeo & Scott,” 2022.
- Kaplan, Jacob, *Uniform Crime Reporting (UCR) Program Data: A Practitioner’s Guide* 2022.
- Kirk, David S., “A natural experiment on residential change and recidivism: Lessons from Hurricane Katrina,” *American Sociological Review*, 2009, 74 (3), 484–505.
- Law Enforcement Support Section and Crime Statistics Management Unit, “Summary Reporting System (SRS) User Manual,” Technical Report 2013.
- Maltz, Michael D. and Joseph Targonski, “A Note on the Use of County-Level UCR Data,” *Journal of Quantitative Criminology*, 2002, 18 (3), 297–318.
- Mueller-Smith, Michael and Kevin T. Schnepel, “Diversion in the Criminal Justice System,” *Review of Economic Studies*, 2021, 88 (2), 883–936.
- Pridemore, William Alex, “A cautionary note on using county-level crime and homicide data,” *Homicide Studies*, 2005, 9 (3), 256–268.
- Ryley, Sarah, Jeremy Singer-Vine, and Sean Campbell, “5 Things to Know About Cities’ Failure to Arrest Shooters,” 2019.
- Savolainen, Jukka and Matthew VanEseltine, “Replication and Research Integrity in Criminology: Introduction to the Special Issue,” *Journal of Contemporary Criminal Justice*, 2018, 34 (3), 236–244.
- United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics., “Law Enforcement Agency Roster (LEAR), 2016,” 2017.
- U.S. Department of Justice Federal Bureau of Investigation, “Uniform Crime Reporting Handbook,” Technical Report 2004.
- Wiltshire, Justin C, “allsynth: (Stacked) Synthetic Control Bias-Correction Utilities for Stata,” 2022.

## A Sample Selection

In Section 2.1 we documented how the main results were not robust to extending the first year in the pre-treatment period. Another selection decision relates to the agencies that constitute the donor pool. As we discuss in detail in this article, Hogan (2022a) fails to account for differences in any dimension other than total homicides, therefore using the top 100 agencies (whose populations range from approximately 300,000 to over 8,500,000) may be a poor comparison for Philadelphia which ranks as the 5<sup>th</sup> largest jurisdiction in the sample. As we did for the pre-treatment start date we systematically vary the cut off (based on average population from 2010-2019) from the top 5 to top 50 jurisdictions.<sup>18</sup> In Figure 7 we present the DiD estimate restricting the donor pool to the top N jurisdictions. Again, if one restricts to donors that are of a similar population to Philadelphia the DiD shrinks considerably and is not distinguishable from 0 at any conventional level of statistical significance.

As a way to visually demonstrate how the donor cities were chosen based on homicide count, Figure 8 shows a scatterplot of homicide counts and rates for every city in Hogan (2022a), averaged across the pre-period (2010-2014) using data from Offenses Known.<sup>19</sup> Philadelphia is shown in the solid green square at 292.6 homicides with a homicide rate of 19.0 homicides per 100,000 people. We also show each of the donor cities—Detroit, New Orleans, and New York City. All other cities are presented as solid black dots. We highlight that the synthetic control unit is comprised of mainly Detroit and New Orleans (80% of the total weight), both cities that have over twice the murders per capita. We believe this substantial discrepancy after accounting for population calls into question the validity of the synthetic unit being a valid counterfactual.

---

<sup>18</sup>There is no variation in the point estimate when expanding from the top 50-500 as donors remain constant.

<sup>19</sup>As Hogan (2022a) notes, Florida data is not available in SHR. It is, however, available in the Offenses Known data.

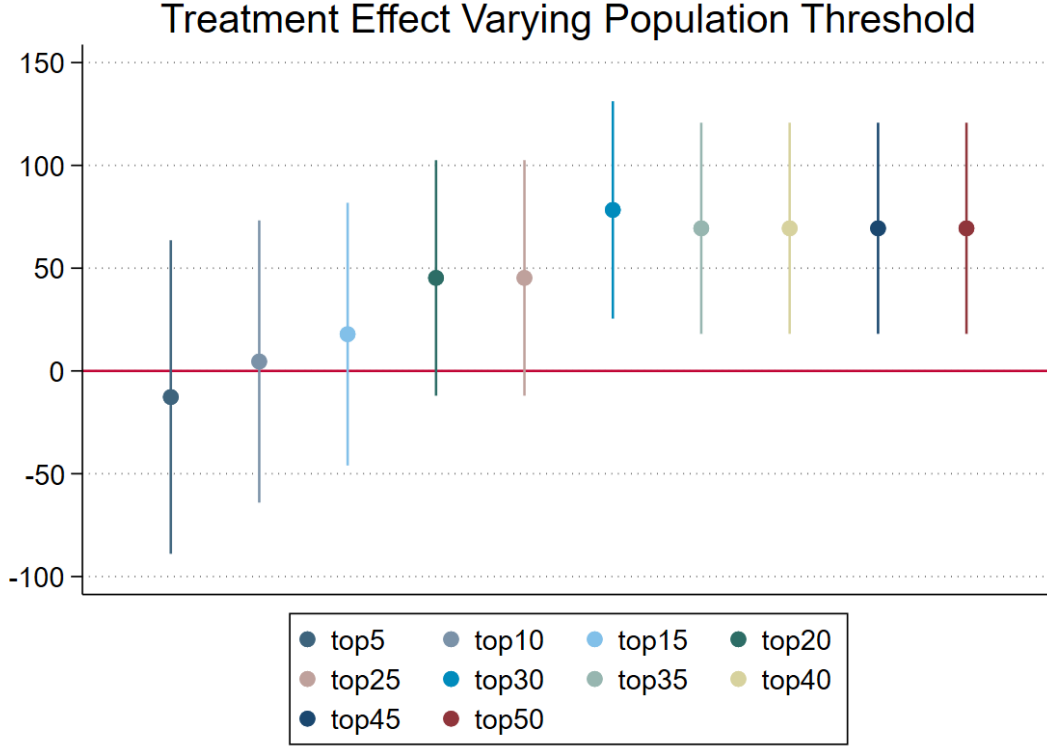


Figure 7: DiD estimates using same data and SCM method in [Hogan \(2022a\)](#) but varying the population cut off of donor pool. Confidence intervals are at the 95% level.

## B Mechanisms and competing theories

An important part in any empirical study of what effect a policy has is to rule out competing explanations and explain how the mechanisms behind the proposed relationship works. One way [Hogan \(2022a\)](#) implicitly does this is by calling Philadelphia’s de-prosecution experience a “natural experiment.” However, this is an incorrect description. A natural experiment involves an event that results in the random assignment of some experience outside of human manipulation. An example is a hurricane that destroys some areas and not others due to randomness in the trajectory of the hurricane ([Kirk, 2009](#)). As [Hogan \(2022a\)](#) clearly states in his manuscript, de-prosecution in Philadelphia was a purposeful policy driven by perceived voter preferences and therefore not randomly assigned to Philadelphia. Because de-prosecution was not randomly assigned, the number and type of confounding factors that



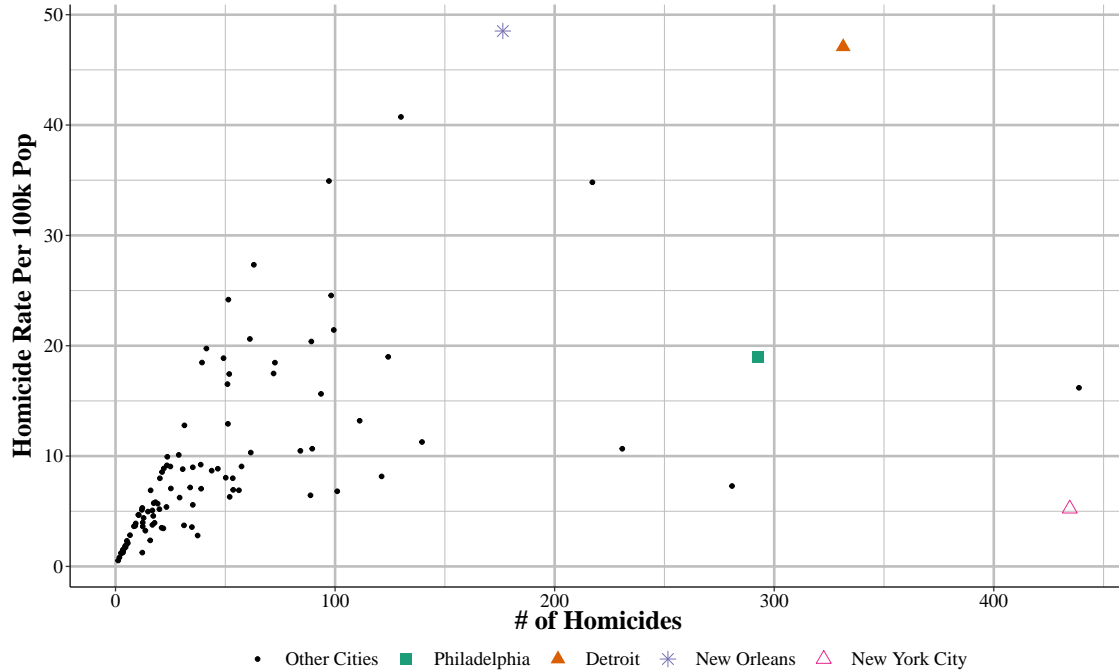


Figure 8: Homicide count vs. homicide rate, averaged across 2010-2014 for agencies included in Hogan (2022), from Offenses Known data

could explain Philadelphia’s change in homicide between 2010-14 and 2015-19 increases. Another way Hogan (2022a) argues for causality is by considering several “competing theories of causation” in the Discussion section of the paper and ruling out each in turn.

One of Hogan (2022a)’s proposed mechanisms behind the de-prosecution and homicides is that de-prosecution decreases the likelihood of an arrest to be made, leading to a cycle of retaliatory murders. The reduction in drug prosecutions, he argues, makes it harder to make arrests as police have less evidence of criminal behavior, which they normally acquired through drug prosecutions, and witnesses are less likely to cooperate in murder investigations. As evidence of this he says that “the closure rate for the Philadelphia Police Department regarding homicides has been declining rapidly during the de-prosecution period.” As a footnote to this sentence he says that “While the Philadelphia Police Department is publicly reporting clearance rates around 50%, the actual underlying SHR data show that clearance rates for 2018–2019 were in the 20%–30% range.” It is unclear where Hogan (2022a) gets these numbers from. It cannot be from the SHR, which does not contain clearance numbers.

It also cannot be from the Philadelphia Police Department’s website, which is Hogan’s source for other Philadelphia murder data, as this site also does not contain clearance numbers. The most likely source is the FBI’s Offenses Known data which [Hogan \(2022a\)](#) uses in the paper.

This data has both the annual number of homicides and the number cleared by the police. Figure 9, which we created using the Offenses Known data, shows the clearance rate for homicides in Philadelphia from 2000 through 2018.<sup>20</sup> [Hogan \(2022a\)](#) claims that clearance rates declined during the post-period, reaching 20-30% by the end of this period. At no point in the studied period did the reported clearance rate reach below 40%. Even the claim that the clearance rate declined during the post-period is incomplete. It certainly did decline, with every year lower than the last other than 2019 (not shown) which contains incomplete data and therefore may be disregarded as not comparable to complete data years. This trend, however, started in 2014, prior to the de-prosecution period. We extend this data through 2000 to demonstrate a second point. The pre-period of 2010-2014 is also a time of declining clearance rates—though less consistent than the post-period—with an increase in 2012 followed by nearly identical rate in 2013 before falling again in 2014. The entire time period of 2000-2019 is a period where rates decline more than they improve.

Perhaps [Hogan \(2022a\)](#) is using whether there is any known demographic information about an offender as a proxy for clearance, as has been done in past research ([Avdiya et al., 2022](#); [Ryley et al., 2019](#)). This approach uses the share of offenders where one of the four demographic traits included in SHR—race, ethnicity, age, and sex—are known. To be clear, this does not mean that who committed the crime is known or whether they were arrested, merely that their demographics is not unknown. For example, if someone was murdered and a witness said that the murderer was a White man, this case would be cleared using this SHR rule. But using the standard in the Offenses Known data, it would only be cleared if at least one arrest was made or the case was closed through exceptional means. When cases are closed

---

<sup>20</sup>We exclude 2019 because that year contains incomplete data for the PPD and therefore may be disregarded as not comparable to complete data years. Using that incomplete data, 66.9% of homicides were cleared in 2019, far from the 20%-30% range cited in [Hogan \(2022a\)](#).

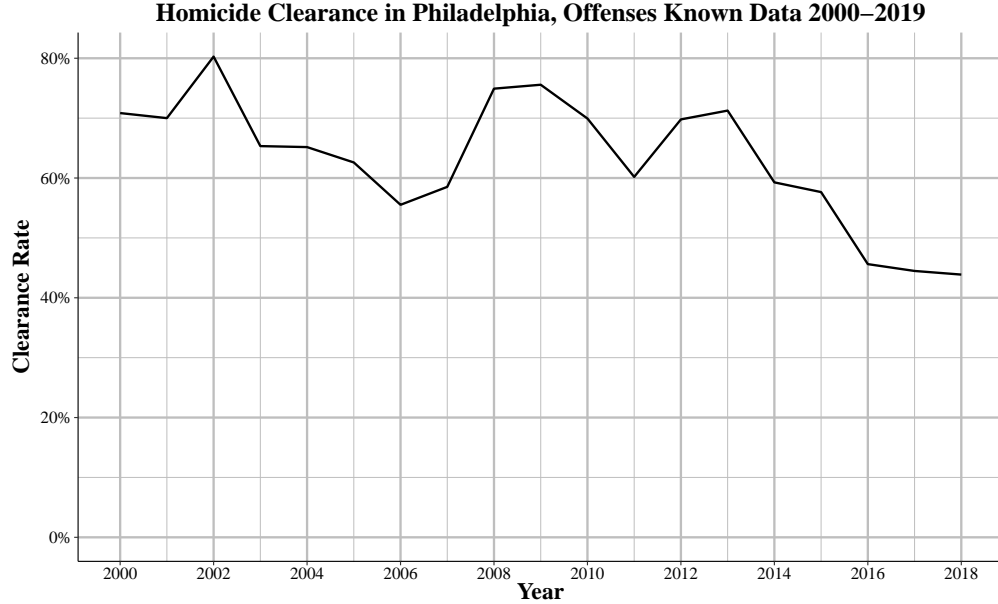


Figure 9: Homicide Clearance in Philadelphia, Offenses Known Data 2000-2019

through exceptional means the police still must have “definitely established the identity of the offender” ([Law Enforcement Support Section and Crime Statistics Management Unit, 2013](#)) but be unable to arrest them for some reason outside of their control.

In Figure 10 we show the percent of incidents where demographic information for an offender is known for all homicides in Philadelphia from 2010 through 2018.<sup>21</sup> For simplicity we examine only the first offender in an incident, even if there are multiple offenders. The data from this method does match the “20-30%” by 2018-2019 as the percent of incidents cleared are 27.35% and 30.35% in these years, respectively. Note that the four demographic variables converge in 2016 where if any are unknown, all are reported unknown. This is not an error in the graph. Accepting these numbers at face value demonstrates a substantial problem in Hogan’s claim: the decline began in the pre-period and clearances declined much faster in the pre-period than in the post-period. Using Sex as an example, which is the best reported demographic variable, the share of incidents cleared decreased from 93.31% in 2010 to 42.52% in 2014, a 54.43% decrease. Comparably, the post-period did much better,

<sup>21</sup>Results are nearly identical when limiting the data to murders and when including 2019, which we exclude as there is only partial-year data available for that year.

dropping from 43.11% in 2015 to 31.68% in 2019, a 26.51% decrease.

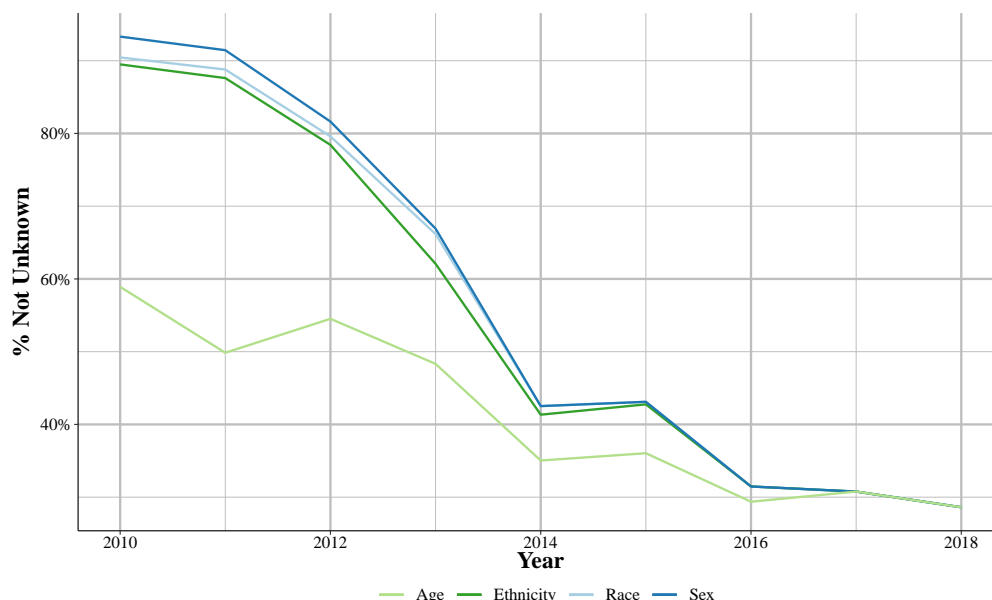


Figure 10: Homicide Clearance Rate Using Known Demographics of Offender, SHR data

A related potential alternative explanation is that de-policing caused the increase in murders, which he describes as the “so-called ‘Ferguson Effect.’” If police are less active then offenders may respond by increasing their criminal behavior, thus leading to an increase in crimes such as murder. He rejects this alternative promptly, saying that “it would apply in a uniform fashion to the large cities in the United States, and thus arguably is not a confounding variable when testing the 100 largest cities during the 2015–2019 post-period.” And that if de-policing is actually only found in “cities where de-prosecution is taking place” then the police response is actually only “a feedback loop caused by de-prosecution.” This is incorrect even at face value. A trend being national does not mean that it impacted each city equally. Nor does he provide evidence that de-policing is either only found only in cities with de-prosecution or that it is in fact caused by de-prosecution.

He argues that de-policing has not occurred in Philadelphia through his Figure 11, shown below in Figure 11. This graph shows the annual number of crimes reported to the Philadelphia Police Department, for each year in the studied period. Based solely on this figure, which he says measures police activity, he claims that there is no evidence of

de-policing. “Figure 11 shows little difference in police activity year-over-year within each crime category. Meanwhile, the number of homicides changes drastically, as homicides are responding to other forces (proposed in this article to be de-prosecution) ... Overall, the data in Figure 11 suggest that the change in behavior in Philadelphia is not with the police department, but with the Philadelphia DAO.”

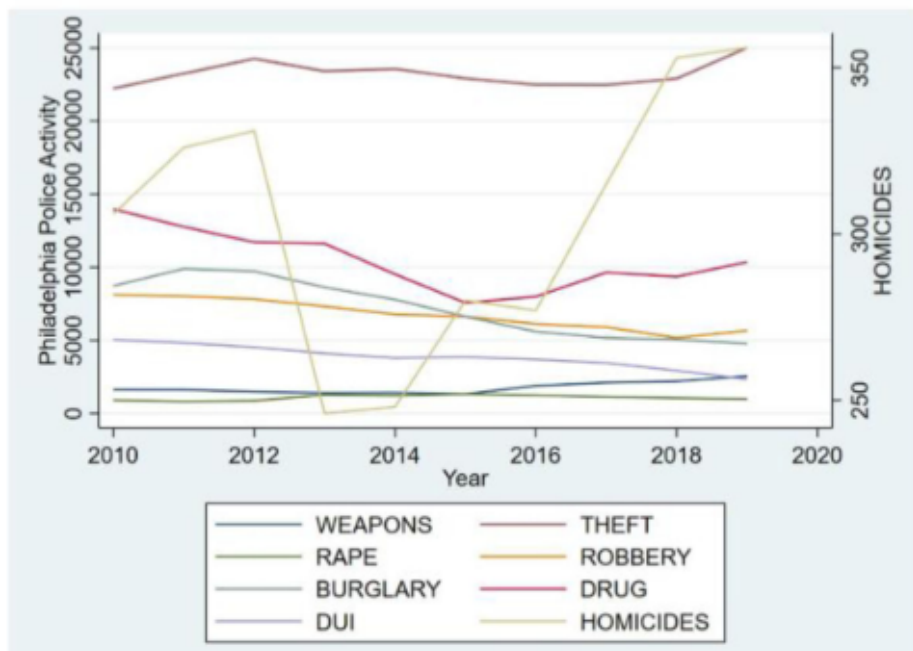


Figure 11: Figure 11 from [Hogan \(2022a\)](#)

With the exception of weapon offenses, whose crime counts he lists for every year, the entirety of the analysis of this data is that the line showing homicide goes up more than the lines for other crime categories. This is certainly true. Every other crime moves very little over the studied period; only homicide shows major swings. However, this is merely an artifact of that non-homicide crimes are on a single scale from 0 to 25,000 while homicide is on its own scale from 250 to 350. Even enormous swings in crime counts would appear to move very little when on a scale that dwarfs the number of crimes that ever occur. We make no claim that [Hogan \(2022a\)](#) intended to mislead readers with this graph, merely that the result is a misleading figure that exaggerates changes in homicide while minimizing changes

in every other crime. To demonstrate this, we recreate the figure and show each crime on its own scale. We do this in two ways: first, using the default scale for the software that generates the graph (for us, the R programming language) and second by forcing the scale to start at 0, as [Hogan \(2022a\)](#) does for all non-homicide crimes.

As shown in Figure 12—contrary to Hogan’s claim—there are seemingly large changes in every crime category over the studied period. For example, burglary drops nearly in half, drugs drop by over 40% before increasing in 2016. Even theft, which ranged from 22,234 crimes in 2010 to 25,013 in 2019, a comparably small 12.5% increase, appears to have massive swings when using the default y-axis range.

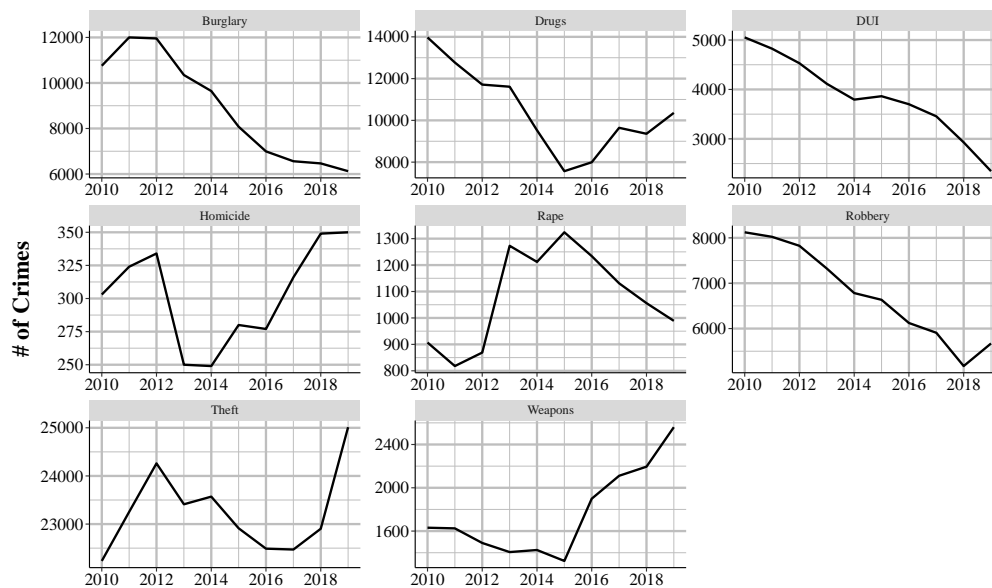


Figure 12: Figure 11 from [Hogan \(2022a\)](#) but separated by crime type and re-scaled y-axis

There is much less movement when forcing the scale to start at zero, as shown in Figure 13. Even homicides show a much more muted movement. Using the default scale makes changes appear larger than they are; forcing the y-axis to start at zero makes them appear smaller. Using them together, as [Hogan \(2022a\)](#)’s Figure 11 does without explanation, misleadingly makes homicide appear to greatly vary while all other crimes remain relatively stagnant.

It is also unclear why [Hogan \(2022a\)](#) defines “police activity” as crimes. An alternative,

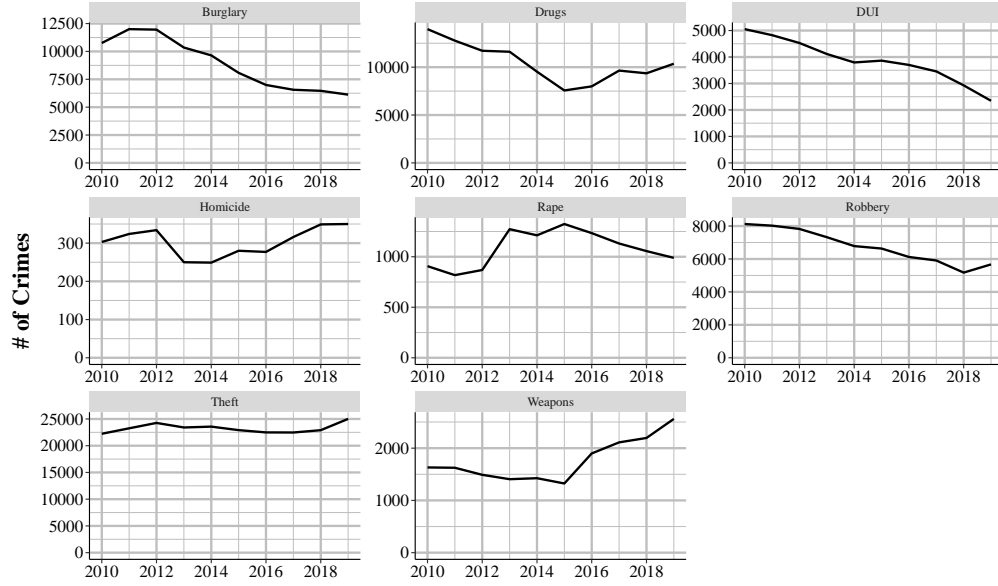


Figure 13: Figure 11 from [Hogan \(2022a\)](#) but separated by crime type with y-axis forced to start at zero

and more commonly used way, to measure police activity is to measure the number of arrests. This is certainly an imprecise way as arrests (like prosecutions) are affected by a number of factors outside of police (or prosecutor) control such as changing laws or true changes in crime. We present a basic example of how total arrests changed over the study period using data from the FBI’s Arrests by Age, Sex, and Race dataset. In Table 2 we show a simple example of arrest trends with the annual number of arrests (for any offense) in Philadelphia. We then replicate columns from Table 2 of [Hogan \(2022a\)](#), showing the number of prosecutions, sentencings, and homicides. We then divide the number of prosecutions and sentencings by arrests to get the share of arrests leading to each outcome (we present this as a percent rather than a proportion for easier reading). Note that we do not include 2019 data as the Philadelphia Police Department only reported arrests for parts of the year and should not be compared with full-year data for each of the other columns.

We include Table 2 both to show an alternative measure of police activity and to demonstrate that change in the units of analysis can drastically alter interpretation of [Hogan \(2022a\)](#)’s result. Looking at the Arrests column we can see that arrests are relatively steady

Table 2: Arrests, Prosecutions, and Sentencings (Table 2 in [Hogan \(2022a\)](#))

Year	Arrests	New Prosecutions	Sentencings	$\frac{\text{Prosecutions}}{\text{Arrests}}$	$\frac{\text{Sentencings}}{\text{Arrests}}$	Homicides
2010	89,692	16,000	6,230	17.84	6.95	306
2011	82,122	14,702	5,147	17.9	6.27	326
2012	86,742	15,334	7,308	17.68	8.42	331
2013	85,954	15,743	6,953	18.32	8.09	246
2014	84,525	14,401	7,252	17.04	8.58	248
2015	69,131	13,140	4,688	19.01	6.78	280
2016	53,725	11,789	5,986	21.94	11.14	277
2017	52,615	11,034	4,423	20.97	8.41	315
2018	46,230	9,036	3,609	19.55	7.81	353

from 2010 through 2014 before dropping substantially. From 2014 to 2018 there is a 45.31% decline in arrests, more than the decline in prosecutions (37.25%) but less than the decline in sentencings (50.23%). This suggests substantial de-policing, which at various points in the paper [Hogan \(2022a\)](#) says is merely “a feedback loop” caused by de-prosecution. We make no claim as to the primary cause of the decline in arrests other than that this is an important empirical question and one that cannot be dismissed without evidence.

As arrests declined faster than prosecutions, the Prosecutions/Arrests column in Table 2 shows that the post-period actually has a higher rate of prosecution per arrest than in the pre-period. Similarly, sentencings per arrest peaked during the post-period and on average have a higher rate of sentencings than the pre-period with 8.54% and 7.66%, respectively. When defining de-prosecution as the share of cases prosecuted, Hogan’s results can be re-interpreted as showing the more prosecution leads to significantly more homicides.

We readily admit that this is a flawed and overly simplistic way of defining de-prosecution. And indeed it may be true that a feedback loop caused police to make fewer arrests as they knew these offenses wouldn’t be prosecuted, therefore the decline in arrests is merely a signal of the decline in prosecutions, not a cause. [Hogan \(2022a\)](#) argues that starting in 2015 not only did *something* change but de-prosecution—and only de-prosecution—changed, causing the increase in homicides. That a simple change from total prosecutions/sentencings to adjusting for arrests can change interpretation in the results so drastically demonstrates that



Hogan's findings require far more thorough testing to rule out possible alternatives.

## C The Black Box of “De-prosecution and death” and the Importance of Reproducibility in Science

On August 23, 2022, Thomas Hogan posted a reply to our response (in the main text of this file) to his initial published manuscript to his SubStack and submitted it for publication in *Criminology & Public Policy*.<sup>22</sup> We refer to Mr. Hogan’s initial manuscript as [Hogan \(2022a\)](#), our response (in the main text of this file) as Kaplan et al. (2022), and his recent reply, entitled “DE-PROSECUTION AND DEATH: A CORDIAL REPLY TO KAPLAN, NADDEO & SCOTT”, as [Hogan \(2022b\)](#). [Hogan \(2022b\)](#) states there are major faults in Kaplan et al. (2022) and that our response actually provides further support for his original methods and findings. Specifically, his primary issue with Kaplan et al. (2022) is that it is “based on critical and obvious errors in data” that when corrected supports his findings. In this brief note, we explain that although one of our analyses uses different clearance data than [Hogan \(2022a\)](#)—a function of [Hogan \(2022a\)](#) 1) not explaining his data and methods, 2) not showing descriptive statistics, or 3) not providing data or code to support an independent reproduction—the other issues we raised and echo here support our previous statements regarding the fatal flaws in [Hogan \(2022a\)](#)’s analyses. We also attempt to bring our clearance data closer to those used in [Hogan \(2022a\)](#), within the constraint of knowing little about his measure, and find that the results are sensitive to if and how his control variables are modeled, and, under certain model decisions, remain nullified.

We repeat our call to Mr. Hogan to share his code and data, at least privately, for the purpose of reproduction. The claim in [Hogan \(2022b\)](#) that senior academics advised him that “researchers usually do not share their data for ongoing research” is a poor, regressive perspective for senior scholars to hold, especially when the request is for the sole purpose of reproduction, the data are secondary and publicly available, and one of the requesting authors created the untransformed data file used in the study. [Hogan \(2022b\)](#)’s other claim

---

<sup>22</sup>To be specific, his post is from his SubStack available [here](#), not to be confused with his other SubStack, available [here](#).

that independent replications would be “biased by shared data or code” is just silly.

## C.1 Controlling for Clearance Rates

Regarding the clearance data used in our response, it seems that we used different data to measure homicide clearances than what was used in [Hogan \(2022a\)](#).<sup>23</sup> We used the raw UCR Offenses Known and Clearances by Arrest data to measure clearance while it appears from [Hogan \(2022b\)](#) that [Hogan \(2022a\)](#) used some undescribed method<sup>24</sup> to modify the UCR Supplementary Homicide Reports (SHR) data to create a proxy for clearance.<sup>25</sup> Importantly, nowhere in his original article did [Hogan \(2022a\)](#) mention where he obtained his homicide clearance count or rate measures or if/how the data were transformed, and he did not share his data or code with us following multiple requests. We also note that elsewhere in his paper, [Hogan \(2022a\)](#) used Offenses Known and Clearances by Arrest data, both for additional crimes studied (he examined burglary and robbery) and even for his primary outcome of the number of homicides for at least one analysis. As we demonstrated in Section 3.1 of our response, [Hogan \(2022a\)](#) used Offenses Known and Clearances by Arrest data, not SHR data, when performing a robustness check studying Baltimore rather than Philadelphia.<sup>26</sup>

For these reasons and because it is the predominant method for measuring clearances

---

<sup>23</sup>It is comedic that Mr. Hogan identified this after reviewing our publicly provided data when he has not given anyone the opportunity to review the accuracy of his data.

<sup>24</sup>[Hogan \(2022a\)](#) and [Hogan \(2022b\)](#) do not explain how he transformed his clearance data, and, based on footnote 8 in [Hogan \(2022b\)](#), it seems that at least for New York City he used a unique, custom imputation method.

<sup>25</sup>The SHR does not measure whether a homicide was cleared. According to [Hogan \(2022b\)](#), the author created a proxy to measure homicide clearance in [Hogan \(2022a\)](#) based on whether the SHR recorded suspect demographic information regardless of whether an arrest was made or whether police knew the identity of the offender beyond knowing one or more of their demographic attributes. Which demographic characteristics had to be present is not explained. While this has been used as a proxy for case clearance in a few publications, it is not a valid measure of case clearance and we do not recommend using it. In this case, there was no need to proxy clearance since the Offenses Known and Clearances by Arrest data have the real measure.

<sup>26</sup>The other city he examined was Chicago, but we could not see if Chicago had the same issue as Baltimore as the figure for Chicago, Figure 9, was identical to the figure for Baltimore, Figure 10, at the time we wrote our response. This has since been corrected, although without any note on the journal or article page, we are unsure when it was changed. It is unclear what the source of the new Chicago data is. In the figure in the current version of [Hogan \(2022a\)](#) the number of homicides in 2019 for “Real Chicago” appears to be greater than 500. According to SHR data, that number is 498 and per Offenses Known and Clearances by Arrest data, that number is 492.

using the Federal Bureau of Investigation (FBI) data and the way the FBI reports clearances in both its annual Crime in the United States report and its official crime data website, we incorrectly assumed that [Hogan \(2022a\)](#) used the raw clearance counts and rates obtained from the Offenses Known and Clearances by Arrest data. Although we admit that we should have taken a closer look at the clearance data, our purpose was to reproduce [Hogan \(2022a\)](#)’s analyses, not to conduct new analyses using different data, and we were going off of the information available to us in [Hogan \(2022a\)](#) about what data were used given the author’s lack of response to our emails. Not only would Mr. Hogan sharing his data/code with us, even under strict use and sharing conditions, have saved us time and effort during our reproduction, it would have allowed for an accurate reproduction, which should be a priority for anyone interested in advancing a scientific evidence base to guide policy. We hope that this interaction serves as a cautionary tale for other researchers about the importance of open data and code for supporting accurate reproductions to improve scientific knowledge.

We again attempt to peer into the black box that is Mr. Hogan’s data, methods, and analyses to bring our clearance data closer to that used in [Hogan \(2022b\)](#). In the abstract of [Hogan \(2022b\)](#), the New York City Police Department’s 0% clearance rates for years 2010-2012 are held as examples of “obvious instances of incorrect data”. Therefore, we develop a systematic method to replace any instance where a large locality reports 0 cleared murders in our data. To do this, we replace all observations that have more than 100 homicides recorded in the UCR Offenses Known and Clearances by Arrest data and 0 cleared homicides with the mean of homicides cleared within the agency for all non-zero observations in 2010-2019.<sup>27</sup> We also note here that we do not make the effort to merge in median income estimates without the author’s data or code. We doubt that this greatly impacts our results and if anything provides another imbalanced predictor which potentially would drive a further wedge between the traditional and bias-corrected results.

---

<sup>27</sup>Our results are not substantively different if instead we use the median and/or if we vary the 100 homicide definition from 50-150. Ideally, we would be able to use Mr. Hogan’s imputation method to directly test his claims; however, he does not describe these methods and did not provide his data/code.

Proceeding to the main analysis, we re-run the bias correction analysis of homicide counts with the imputed clearance data and again find that the DiD estimate is not robust to correcting in a linear way for the potential bias induced by not taking into account pre-period differences in predictor variables (e.g., population and clearance rates). Again, because Hogan (2022b) did not share his code or adequately describe his methods, we cannot know exactly what model he ran in his updated bias correction analysis. As we described in Kaplan et al. (2022), and can be seen in our publicly posted code, we used a linear specification when correcting for bias. Here with the more complete clearance data, we re-run our model using ridge, lasso, elastic net, and a linear model that only uses donors with positive weight to estimate the correction equation.<sup>28</sup> Figure 14 presents the point estimate and 95% confidence interval for each regression method.

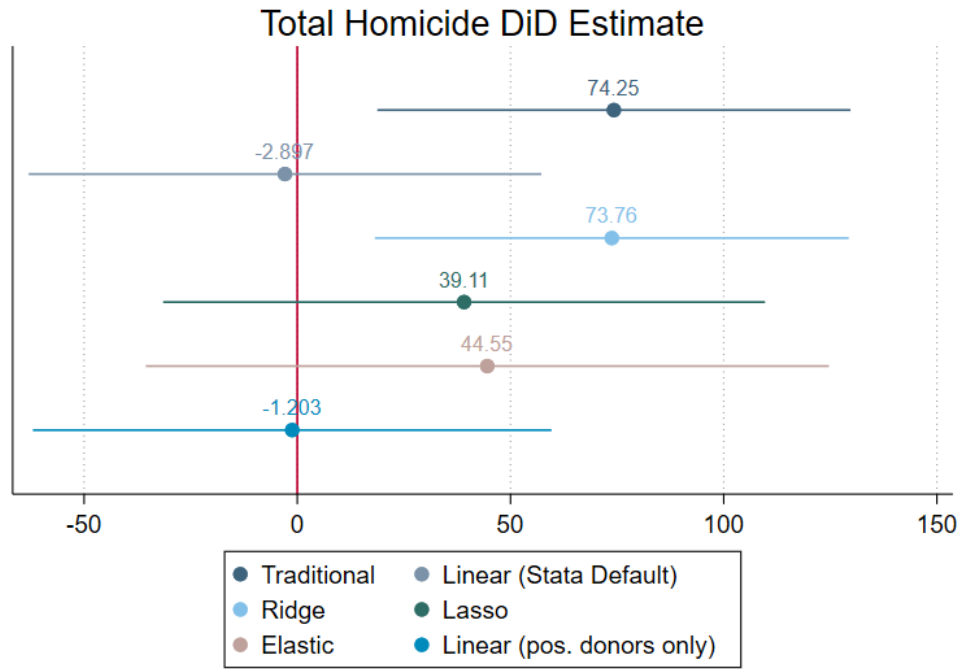


Figure 14: Difference in differences estimate using traditional and bias-corrected synthetic control models, with different models used to adjust for bias. 95% confidence intervals constructed using robust standard errors.

One can see that there is considerable heterogeneity in the results based on modeling

<sup>28</sup>These are all the options available in *allsynth* package in Stata.

decisions, providing further evidence that the results in [Hogan \(2022a\)](#) and [Hogan \(2022b\)](#) are not robust outside certain analytic specifications. For example, the bias correction procedure produces almost identical results to results from the traditional SCM estimator if ridge regression is used, and slightly negative results if a linear specification is used. In fact, as shown in [Figure 14](#), in 4 out of the 5 available regression models, the main results are statistically indistinguishable from zero at the conventional 95% level. Again, we note the consistent themes of only those results that support the author’s hypotheses showing up in his results and a lack of documentation of modeling decisions.

To reiterate, the objective of our study was not to “reverse p-hack” Mr. Hogan’s results. Instead, it was to reproduce his findings using more justifiable modeling decisions such as using a rate as the outcome, accounting for covariates, and using a longer pre-intervention period, and to test the sensitivity of our results to these specifications. Our reproduction findings led us to the conclusion that the results in [Hogan \(2022a\)](#) appeared to be the product of modeling decisions that 1) were poorly explained and documented (which made replication difficult), 2) were based on unjustifiable model specifications, and 3) maximized the positive effect found. With clearance data that more closely resemble the author’s data, we again find that the author’s results are sensitive to modeling decisions. This is of particular concern given that the more appropriate rate-based outcome did not undergo all of the robustness checks applied to the count-based outcome in [Hogan \(2022a\)](#).<sup>29</sup> Therefore, we stand by our claim that Mr. Hogan’s findings should not be used to inform policy.

## C.2 Remaining Issues

Having discussed the most significant claim in [Hogan \(2022b\)](#) regarding our response, we next respond to several other minor claims that he makes about [Hogan \(2022a\)](#) and [Kaplan et al. \(2022\)](#).

---

<sup>29</sup>As explained in [Kaplan et al. \(2022\)](#), we are not suggesting the author should have interpreted the rate-based model given that it did not identify a unique solution. We are saying that a count-based model is not an appropriate substitution for a rate-based model simply because of this finding.

In paragraph 2 of Section 2, Mr. Hogan claims that our use of Philadelphia Police Department’s (PPD) SHR data was to “challenge the results” of his original article. Rather, it was to explain our process of trying to reproduce the results in the original article, which states that the SHR data were used to measure the outcome, with no statement that the PPD data were used just for the treatment city. We believe that this is clear in our writing. He ends this paragraph with a startling, bizarre, and certainly uncordial accusation: “The Kaplan Response is literally hiding dead bodies, a well-respected tactic in criminal circles, but a disfavored practice in law enforcement and academics.”

In paragraph 3 of Section 2, Mr. Hogan lists a number of issues that we allegedly “concede”, “admit”, or “do not challenge”. Our response had the narrow scope of addressing a few critical errors in the author’s original manuscript. This was due to space limitations and that flaws in the data and analyses stood out to us as more problematic and less subjective than other issues in the article. How Mr. Hogan classified cities based on prosecutor philosophy, what population factors were or were not changing at the time, and whether the synthetic control method is the best, or even an appropriate method for testing de-prosecution in this case were not discussed in our paper. It is not correct to interpret our article’s lack of discussion of certain claims or decisions as an agreement with them.

Relatedly, Mr. Hogan says that we actually support his incorrect measure of crime clearance. He states, “Regarding the clearance data, the Kaplan Response correctly ascertained that the De-Prosecution Article relied upon demographic information for offenders to calculate the clearance data, a practice that the Kaplan Response explicitly notes is acceptable, and eventually concedes that the data are accurate as calculated in the De-Prosecution Article (Kaplan et al., 2022; Avdija et al., 2021).” We reiterate that Mr. Hogan does not actually measure whether a case was cleared, merely whether an offender’s demographic trait—race, gender, age, or ethnicity—was reported in the FBI’s SHR data. In the case of a murderer who police know is male but have no idea about his identity, the case would be considered cleared under Hogan’s definition, but not according to law enforcement or

the FBI. The fact that other authors use this same inaccurate measure makes it neither acceptable nor correct.

At the end of Section 3.1, [Hogan \(2022b\)](#) states “In this case, the augmented synthetic control method is not called for, being more likely to confound results than reduce bias.” We are unaware of any case in which controlling for a basic set of theoretical confounders would lead to more biased results than only controlling for lagged values of an outcome. In this case, because [Hogan \(2022a\)](#) used a short pre-intervention period and did not account for the impact of population (or any other predictor) on homicide counts, we maintain our view that bias correction methods should be used to test the robustness of the original SCM results. Additionally, [Hogan \(2022b\)](#), at the end of Section 3.1, justifies his methods by comparing his study with [Abadie et al. \(2010\)](#) and noting that the authors of that article “... did not see the need to modify the traditional synthetic control algorithm in this context with bias adjustments, such as the later conceived augmented synthetic controls.” This is an illogical claim for three reasons.

First, and perhaps most obvious, the proposed adjustments were published more than a decade later in a series of papers ranging from 2019-2021.<sup>30</sup> Any good statistical method undergoes advancements over time, and the fact that the original SCM was once used when those advancements did not exist does not justify its current use. Second, [Abadie et al. \(2010\)](#) used a per capita rate outcome for the same reason [Hogan \(2022a\)](#) should have used per capita homicide rates as the outcome to account for bias caused by the correlation between population size and a count outcome. Third, despite [Hogan \(2022b\)](#) claim that the augmented SCM is not called for and is likely to confound the results, he goes on to conduct this model later in his reply as a “robustness test”. Because Mr. Hogan refuses to share his code and data and has not adequately described his methods in either article, one is forced to trust him that his new results are accurate. Importantly, because the figure appears to be

---

<sup>30</sup>The authors concede that while economics is notorious for a lengthy publication process, it was likely that the bias-correction methods ([Abadie and L’Hour, 2021](#)) were not yet conceived (or at least well vetted) at the time of the [Abadie et al. \(2010\)](#) paper. Or, more likely, the 2010 paper was written in 2000 and the 2020 paper was written in 2010.



an exact replica of the original finding and greatly differs from the results we obtain (even after adjusting for more accurate clearance rates), we are left doubting the validity of this new analysis.

The first paragraph of Section 3.2 confuses the interpretation of SCM findings in both Kaplan et al. (2022) and Hogan (2022a). Hogan (2022b) first cites the estimate of -46.22 presented in Kaplan et al. (2022) and then states:

To engage the illogical results proposed by the Kaplan Response, starting from a baseline of 248 homicides in 2014 and converting their estimating model into a predictor, their model predicts that Philadelphia should have recorded only 18 homicides in 2020 and none (or -28 homicides) in 2021.

In fact, our results do not imply that Philadelphia should have recorded only 18 homicides in 2020 and -28 in 2021 but that on average Philadelphia is estimated to have roughly 46 fewer homicides compared to its synthetic control. Therefore, to back out a level from our estimates, one must first calculate the total number of homicides the synthetic control experienced in 2020 and 2021 and subtract off 46. This exemplifies a recurring trend, whereby Mr. Hogan either genuinely does not understand properties of his models/estimates and/or knowingly manipulates their interpretation to create statistics that fit a narrative. Additionally, Mr. Hogan refers to our “heavily manipulated model”, but does not give a basis for this claim. It seems to us that he is mistaking manipulation with bias correction. Finally, we agree with the premise of the argument that if results are dramatically large and/or dissimilar to extant literature, applied researchers should take this as “a strong signal that their methodology [may be] seriously flawed” (from Section 3.2 of Hogan (2022b) with insertion by authors). This was in fact the reason we set out to understand the original results in Hogan (2022a), as they 1) represent a massive effect of an approximately 31% increase in total homicides (compared to the mean homicide count of 242 for the synthetic control from 2015-2019) and 2) diverge from quality research that shows modest decreases

in crime seemingly caused by de-prosecution.<sup>31</sup>

In footnote 5, [Hogan \(2022b\)](#) says “The Kaplan Response professes not to understand how New York City, composed of five boroughs, was classified as ‘middle’. Three of the boroughs were identified as ‘traditional’ or ‘middle’, while two of the boroughs were identified as ‘progressive’, yielding an aggregate classification of ‘middle’.” We certainly did not understand how New York City was classified as nowhere in the original paper did he explain this. This reiterates the importance of providing either a detailed methodology section and/or one’s code to adequately communicate scientific methods and findings.

In the last paragraph of Section 4.1, Mr. Hogan states that Kaplan et al. (2022) took the position that the SHR data cannot be trusted. This is not true. Our criticism was that by not describing his procedure for cleaning the data and imputing missing data, including how often data were imputed and from where, one must be skeptical of how comparable the different homicide data sources used in [Hogan \(2022a\)](#) are to each other. For example, the homicide counts obtained from the Philadelphia Police Department are homicide victim counts, while other agencies may only report homicide incident counts.

Section 4.2 does not address our criticisms, so we do not offer a response. We leave it to the reader to decide whether they are content with a count outcome.

For Section 4.3, we stand by our criticism that [Hogan \(2022a\)](#) use of a short pre-intervention period likely resulted in biased analyses. The purpose of using a long pre-intervention period is not to test whether the results vary over time, as suggested in [Hogan \(2022b\)](#), but to ensure that the treated and synthetic control units are truly equivalent. If Philadelphia and Synthetic Philadelphia were equivalent before 2015, which is the major assumption behind using the SCM to test for a causal impact of the start of de-prosecution during that year on homicide, it is unclear why the two units would not respond similarly

---

<sup>31</sup>See [Agan et al. \(2021\)](#) and [Mueller-Smith and Schnepel \(2021\)](#) for examples. To be clear, we believe that results that differ from existing research are important and potentially more interesting, as they elucidate nuances in the topic being studied. However, we believe that when these results are discovered, there is a burden to at least discuss potential mechanisms and/or differences in settings that could explain why results differ.

to prior events like the Great Recession. If there are factors about the two units that would cause them to respond differently to the prior event, this suggests to us that the two units are not equivalent and that the post-2014 difference in trajectories could be caused by these pre-existing differences and not the difference in prosecution practices. This is the interpolation bias [Abadie \(2021\)](#) discusses that is not discussed in [Hogan \(2022a,b\)](#). Simply put, in SCM analyses a relatively long pre-intervention period gives you greater confidence that all else is actually—not just seemingly during a short period—being held equal.

Finally, [Hogan \(2022b\)](#) references the [Abadie et al. \(2010\)](#) study to provide evidence that the population size differences among his sample of large cities are not problematic. In fact, the differences in population size between the states used in [Abadie et al. \(2010\)](#) were not an issue because those authors used a per capita rate as the outcome. Had they simply used the number of cigarettes sold as their outcome, one would think someone would have called them out for the obvious correlation between the number of people who live in an area and the number of cigarettes purchased. That is why it is important to use per-capita rates.

### C.3 Conclusion

Since publishing [Kaplan et al. \(2022\)](#), several people, including [Hogan \(2022b\)](#), have questioned our personal opinions on de-prosecution, implying that we attempted to reproduce [Hogan \(2022a\)](#) results and pointed out the flaws of the study because the findings were contrary to our beliefs about prosecution or crime control. To us, this is a reflection of how rare scientific reproduction and replication is in criminology that the few of them that are completed appear to be motivated by bad intentions. In this case, we read a recently published paper in a well-respected journal and discussed the article with colleagues, as is typical in any field. The authors happened to agree that there were so many issues with the paper that it deserved a closer look and, ultimately, a formal response. Of course, this answer will not satisfy everyone. In the end, it will be up to the readers to understand the issues discussed in this exchange and come to their own conclusions about the merits of each

paper. We hope that social science associations, journal editors, and research institutions will continue to seriously consider the issues of reproduction/replication, open science, and promoting scientific objectivity so that criminological research can contribute meaningfully to scientific knowledge and evidence-informed policy.