Appendix
# Chicago's Safe Passage Program to Prevent Crime: Is It Worth the Dime?

Stefan Binder

August 22nd, 2018

## Contents

# 1  Introduction

This Appendix belongs to the website `https://binste.github.io/chicago_safepassage_evaluation/`, which gives an introduction to the project, and presents the empirical strategy as well as the results in a non-formal way. The following sections provide more in-depth information on the respective topics.

# 2  McMillen et al. (2017)

This section provides a more detailed description of McMillen et al. (2017)[1], with a focus on context, research questions, and results.

## 2.1  Context

The work by McMillen et al. (2017) builds upon a vast body of crime prevention literature. As described in the introduction of their paper, previous work has shown that the deployment of police forces to crime hotspots can significantly reduce crime in the targeted areas (see Di Tella and Schargrodsky 2004, Klick and Tabarrok 2005, Draca et al. 2011, Weisburd et al. 2009, Braga et al. 2014, Lum and Koper 2014). However, a stronger involvement of the community itself might provide a viable alternative to an increased presence of the police (Krivo 2014). To examine this claim, McMillen et al. (2017) look at the example of Chicago's Safe Passage program, where civilian guards are deployed along designated high-risk school routes.[2]

## 2.2  Crime

McMillen et al. (2017) use various specifications to assess the effect of the program on crime. The main specification is very similar to the one replicated in this project. The only difference lies in the units of observation, which are artificially constructed cells of one eight by one eighth miles instead of census blocks. I decided to replicate their results derived on the basis of census blocks, as for these, the same boundaries as McMillen et al. (2017) used were available online. Their main specification finds a reduction in violent crime of an average of 14% due to the program, but no statistically significant effect on the number of property crimes. There is no evidence of displacement effects into neighboring areas caused by the program. McMillen et al. (2017) furthermore show that the reduction in crime is persistent throughout the implementation period and does not fade out over the years.

The strategy of McMillen et al. (2017) to estimate the effect of the program on crime as well as the results related to violent and property crimes are further described on the website, as well as in sections 4 and 5 of this Appendix.

### 2.2.1  Cost-benefit analysis

To make these findings more tangible for policy makers, McMillen et al. (2017) performed a cost-benefit analysis of the Safe Passage program with respect to the reduction in crime. They estimate that the program accrues benefits of around $100 million a year, based on a commonly used estimate for the willingness to pay to reduce crime. For the school year 2015-2016, the total costs of the program amounted to $17.8 million. McMillen et al. (2017) conclude that Chicago's Safe Passage program therefore represents a cost-efficient option to reduce crime.

## 2.3  Attendance

In addition to crime counts, McMillen et al. (2017) furthermore look at the effect of the implementation of the Safe Passage program on attendance rates at schools. Using a similar difference in differences approach as for the

---

[1]  At the time of submission of this thesis, McMillen et al. (2017) is under "revise and resubmit" at the Journal of Urban Economics.
[2]  For an introduction on the Chicago's Safe Passage program, see the website at `https://binste.github.io/chicago_safepassage_evaluation/`.

crimes, with the change in attendance rate per school as the outcome variable, they find a decline of 2.5 percentage points in the rate of absenteeism. As control group, public schools are used which are not yet part of the program. To address the concerns of the relative similarity between control and treated schools, McMillen et al. (2017) show that the results are robust to only using schools as controls, which are in the same community areas as the treated ones. In another estimation, each treated school is matched to two control schools using propensity score matching based on pre-program attendance, school characteristics, as well as census block group characteristics. The results are again consistent with the baseline estimation.

The results by McMillen et al. (2017) therefore suggest that involving members of the community can be a cost-effective way to persistently curb crime as well as have a positive effect on school attendance, through deterring criminals.[3]

# 3 Data

Following McMillen et al. (2017), for the replication I mostly used data sources which are available online for free. The only exception is the information, on in which year the Safe Passage program was implemented at what school. This data was obtained via a Freedom of Information Act (FOIA) request and is now, together with the other raw data files (with the exception of the crimes dataset, due to its size), available on the GitHub repository of this analysis.[4] More information on the raw data files such as when they were downloaded as well as the corresponding URLs can be found in the Jupyter notebook *notebooks/0_download_data/0.0-bis-download-data.ipynb*.[5]

The data preparation was done in Python using the following packages:[6] numpy (Oliphant 2006), pandas (McKinney et al. 2010), geopandas, altair, matplotlib (Hunter 2007), sqlalchemy, requests, tqdm (da Costa-Luis et al. 2018), shapely (da Costa-Luis et al. 2018), ipywidgets, and Jupyter notebook (Kluyver et al. 2016).

## 3.1 Crimes

The City of Chicago's Police Department provides extensive information on all reported crime incidents since 2001 through the Chicago Data Portal.[7] The dataset includes date and time of the reported occurrence of a crime as well as an address and corresponding GPS coordinates. The information on the location of the crimes are slightly obfuscated to the level of the corresponding census block. However, this still allows to aggregate accurate crime counts on the block level, which will be the level of spatial granularity used in this analysis. For the relevant time period of 2006-01-01 to 2016-06-30, there are 3,736,428 crimes reported. Out of all these, we are only interested in so-called property crimes (1,189,218) and violent crimes (313,250). The categorization in the raw data set is done according to the Illinois Uniform Crime Reporting (IUCR) code, which adheres to the FBI Uniform Crime Reporting (UCR) program.[8] See Table 1 for the definitions of the two categories.

---

[3]  To my knowledge, the only other analysis of the Chicago Safe Passage program was conducted by Ali (2015), and compares crime counts for 64 Safe Passage routes between the school year of 2012-2013 and 2014-2015. He found, that crime decreased between 6am and 6pm on school days by 26% between the two school years. Note that it is not further specified if this refers to all reported crime incidents or specific subcategories. However, their comparison cannot account for other factors which could have caused the decline.

[4]  See `https://github.com/binste/chicago_safepassage_evaluation/tree/master/data/raw`

[5]  See          `https://github.com/binste/chicago_safepassage_evaluation/blob/master/notebooks/0_download_data/0.0-binste-download-data.ipynb`

[6]  This is a non-exhaustive list and dependencies of these packages might not have been considered. Citations are provided were found on any official website of the package.

[7]  `https://data.cityofchicago.org/`

[8]  See `https://ucr.fbi.gov/`, accessed on 5 May 2018.

Table 1: Relevant crime categories

| Crime category | Uniform Crime Reporting (UCR) code and description |
| --- | --- |
| Violent crime | • Homicide 1st & 2nd degree (01A)<br>• Criminal sexual assault (02)<br>• Robbery (03)<br>• Aggravated assault (04A)<br>• Aggravated battery (04B) |
| Property crime | • Burglary (05)<br>• Larceny (06)<br>• Motor vehicle theft (07)<br>• Arson (09) |

Source: `http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html`, accessed on 28 April 2018

As information on the location of the crime is essential for the analysis, crimes with missing GPS coordinates were dropped. This affects a total of 14,661 incidents out of the violent and property crimes. The number of crimes with missing GPS coordinates varies widely over the years. However, in and of itself this does not pose a problem for the chosen estimation strategy. Problematic would be a systematic difference between treatment and control blocks.[9]

Following McMillen et al. (2017), I only consider crimes which took place during school days, i.e. I drop incidents which occurred on a weekend or in the two summer months July and August. Furthermore, I follow the restriction of dropping crimes that happened outside of the presence hours of the guards (before 6.30am or after 5.30pm).[10] After matching the crimes via their location with the census blocks, one obtains a total of 485,363 reported crime incidents.

McMillen et al. (2017) list several limitations of the dataset, such as that the data only contains crimes which are reported to the police and is therefore not an exhaustive list of all crime incidents. This could pose a threat to the internal validity of the estimates if crimes were systematically reported differently in the treated and control blocks, and if this correlated with the implementation of the program. However, there does not seem to be an obvious reason why this should be the case. A further limitation is that the reported time for a crime is approximated on the hour if the exact time is unknown. Reporting errors regarding the exact time of the crime do not pose a threat to the identification strategy used, as crime counts will be aggregated per month.

## 3.2 Implementation of Safe Passage program

The data on the initial school year of the Safe Passage program per school was obtained through a FOIA request to the Chicago Public Schools.[11] The initial request can be found on the official CPS FOIA website[12] by going to the "Archive" and searching for my name. The provided data unfortunately does not cover school years after 2015-2016. This would be the only additional information needed to extend this analysis and study the effect on the school years 2016-2017 onwards.

---

[9]  Some of the crimes with missing GPS coordinates provide an address from which the coordinates could be recovered by using a geocoding web service. As these services are not free of charge for such a large number of observations, it was not done for this analysis. Recovering the missing GPS coordinates would give the ability to check for any patterns which could pose a threat to the internal validity of the estimation.

[10] The exact times are school specific and unknown.

[11] This is presumably the same data as McMillen et al. (2017) obtained for their analysis, as the number of schools per school year is exactly the same as in Table 1 in McMillen et al. (2017). See Table 2 for a summary.

[12] `https://cps.edu/About_CPS/Departments/Law/Pages/FOIARequest.aspx`, accessed on 18 August 2018.

Table 2 summarizes the number of schools added per school year.

Table 2: Rollout of Safe Passage program

| School year | Number of schools added |
|---|:---:|
| 2009-2010 | 35 |
| 2010-2011 | 0 |
| 2011-2012 | 0 |
| 2012-2013 | 4 |
| 2013-2014 | 55 |
| 2014-2015 | 39 |
| 2015-2016 | 7 |
| Total | 140 |

This is an exact replication of Table 1 in McMillen et al. (2017).

In the raw data, the number of schools added prior to the school year 2011-2012 is not contained, and we only know that 35 schools had the program for the school year 2011-2012. For the purpose of this replication, I follow McMillen et al. (2017) in assuming that these 35 schools were already present in the school year 2009-2010. However, this is an assumption which should be further scrutinized.

Data source: Data provided through Freedom of Information Act request by Chicago Public Schools. See Section 3 for more information.

## 3.3 Information on schools and Safe Passage routes

The addresses and corresponding GPS coordinates of the schools were obtained from the Chicago Data Portal.[13]. The information was available for the school years 2013-2014, 2014-2015, and 2015-2016. The location of earlier years was not required for this analysis, as the location of the schools was not used for the final estimation, but only for exploratory analysis and validation of the locations of the Safe Passage routes.

Shapefiles containing characteristics of the routes such as the corresponding school name, as well as the location of the routes in form of GPS coordinates, were also obtained from the Chicago Data Portal. The portal provided the data for the school years 2013-2014, 2014-2015, and 2015-2016. As there was no exact information on the shapes of the routes prior to 2013-2014, the shapes of 2013-2014 were taken as an approximation. Information on which routes were active was obtained by linking the schools belonging to the routes with the data on the implementation of the program obtained through the FOIA request (see Section 3.2). For some schools where the Safe Passage program was implemented in the school years prior to 2015-2016, no information on the routes could be found for the corresponding school year. For these cases, the location of the same route in the school year 2015-2016 was used.

As school names for the same school differed in writing across school years as well as between the school location data and the routes data, I harmonized them manually for the schools which had a designated Safe Passage route

---

[13] See https://data.cityofchicago.org/. For information on when the data was accessed, see Section 3 and the therein referenced notebook.

at least once. This was done using various online news sources and websites of the schools themselves.[14] The other schools are not used in the analysis, and therefore a manual matching was not necessary. Furthermore, schools were checked for changes in addresses, to identify if they have moved or not.[15] A school that changed its location was introduced as a new school at the new location.

## 3.4   2010 census block boundaries

Block boundaries according to the 2010 Census were obtained through the Chicago Data Portal as shapefile. The same block boundaries were then used for all analyzed school years. For the years prior to 2010, it is unknown if the anonymization of the crime locations (see Section 3.1) on a block level was done according to the 2000 Census definition. If block boundaries have changed in between, this would lead to some crimes being attributed to a wrong block before 2010.

# 4   Empirical strategy

This section provides additional information on the regression used by McMillen et al. (2017) – and therefore also in this analysis – to evaluate the effect of the Safe Passage program on crime counts on a census block level. See the website for an introduction.

## 4.1   Model

The effect of the Safe Passage program on violent and property crime counts is estimated using the following Poisson regression:

$$crime\_count_{it} = \beta \; treated\_block_{it} + \gamma_1 \; one\_block\_over_{it} + \gamma_2 \; two\_blocks\_over_{it} + \delta_i + \lambda_t + e_{it}, \qquad (1)$$

where $i$ indexes census blocks and $t$ identifies individual months per school year.[16]  The dependent variable $crime\_count_{it}$ stands for either the monthly violent or property crime count.[17], $treated\_block_{it}$ is an indicator variable taking the value one for blocks in the months that are guarded by Safe Passage personnel, $one\_block\_over_{it}$ and $two\_blocks\_over_{it}$ are indicators for the months after the Safe Passage was enacted and for blocks that are one or two blocks over. $\delta_i$ are block fixed effects to control for time-invariant differences between blocks, and $\lambda_t$ represent time fixed effects to control for time-specific confounders. $e_{it}$ is the error term.

The coefficient of interest is given by $\beta$. If the Safe Passage program reduced crime, we would expect a negative estimate for $\beta$. This specification furthermore allows to analyze potential spillover effects of crimes to neighboring census blocks. Spillover effects would show in a positive estimate of $\gamma_1$ or/and $\gamma_2$. If this were the case, $\beta$ would still give us an unbiased estimate of the causal effect of the program. However, the displacement effects would need to be considered when evaluating the benefit of the crime prevention program. Leaving away the variables $one\_block\_over_{it}$ and $two\_blocks\_over_{it}$ could lead to a downwards-biased estimate of $\beta$ if spillover effects indeed existed, and we would therefore overestimate a possible reduction in crime due to the program. To control for displacement effects in such or similar ways is common in the evaluation of crime prevention policies (see among

---

[14]   The dataset containing the harmonized names as well as all corresponding versions can be found in the GitHub repository, see `https://github.com/binste/chicago_safepassage_evaluation/blob/master/data/raw/school_names.xlsx`.

[15]   A school was identified as changing its location if their was a difference in reported GPS coordinates between school years of more than 500 meters.

[16]   Observations for the months June and August are excluded as they are outside of a school year, see Section 3.1.

[17]   See Section 3.1 for a description of how these counts are calculated. McMillen et al. (2017) use crime counts per fixed unit of land instead of per capita rates, as interest lies in analyzing the effect on the number of crime incidents, instead of on an intensity measure for a given population size. Furthermore, neither are exact estimates of the number of people living in such a small area available, neither do they reflect the actual daytime population and therefore the actual number of people at risk of becoming victim of a crime (Grogger 2002).

others Grogger 2002, Di Tella and Schargrodsky 2004, and Draca et al. 2011).

I follow McMillen et al. (2017) in estimating Equation (1) as a Poisson regression due to the dependent variables (violent or property crime count) being count variables.

The programming language R was used for the estimation of the Poisson regression as well as some minor preprocessing. The relevant packages were glmmML (Broström 2018 as well as tidyverse (Wickham 2017. The Python package rpy2 was used to interface R.

# 5    Main results

Equation (1) was estimated using data from January 2006 to August 2016, following McMillen et al. (2017), for all blocks which had at least once one of the block dummies equal to one.[18] Table 3 presents the results and compares them to the estimation done by McMillen et al. (2017).[19]

Table 3: Estimated coefficients from Poisson regression

|  | *Dependent variable:* | | | |
|---|---|---|---|---|
|  | Violent crime | | Property crime | |
|  | Binder (2018) | McMillen et al. (2017) | Binder (2018) | McMillen et al. (2017) |
| Treated block | -0.134*** | -0.144*** | -0.004 | -0.005 |
|  | (0.025) | (0.036) | (0.013) | (0.025) |
| One block over | 0.010 | 0.005 | 0.012 | 0.026 |
|  | (0.024) | (0.029) | (0.013) | (0.024) |
| Two blocks over | -0.004 | 0.044 | 0.022* | 0.017 |
|  | (0.025) | (0.027) | (0.012) | (0.025) |
| Census block FE | Yes | Yes | Yes | Yes |
| Year-month FE | Yes | Yes | Yes | Yes |
| Observations | 1,085,334 | 783,340 | 1,460,786 | 982,832 |

*P $<$ 0.1; **P $<$ 0.05; ***P $<$ 0.01.
Heteroskedasticity-robust standard errors are reported in parantheses. For Binder (2018), standard errors are clustered at the census block level. McMillen et al. (2017) reported standard errors clustered by Safe Passage routes. See the main text for a discussion.
Data sources for Binder (2018) columns: See Section 3. Estimated for time period of January 2006 to August 2016 for all blocks, which had at least once one of the block dummies equal to one.
The results by McMillen et al. (2017) are taken from Table 10 in the original paper.

When looking at violent crimes as dependent variable, we see that the coefficient of interest, $\beta$ ("Treated block"),

---

[18]   It was not clear from the original paper if McMillen et al. (2017) only included observations up to 5 school years prior to the implementation of a program and 3 school years after the implementation. This restriction would apply for treatment as well as control blocks, whereas for the latter the implementation in the corresponding treatment block would be referenced. However, the results presented in this section are robust to this reduction in the number of observations used. See *notebooks/5_analysis/1.0-binste-analyze-crime-results-census-block-level.ipynb*

[19]   The coefficients of the Poisson regression can be interpreted as $(\exp(\beta) - 1) * 100\%$.

could almost exactly be replicated in terms of magnitude, with a point estimate of $-0.134$ compared to the $-0.144$ from McMillen et al. (2017), as well as statistical significance. Both the original paper as well as the replication did not find any evidence of spillover effects, which would be captured by the coefficients $\gamma_1$ ("One block over") and $\gamma_2$ ("Two blocks over").

The replication was similarly successful when looking at property crime counts as the dependent variable. No significant causal effect of the program could be found, and the point estimates for $\beta$ are almost exactly the same. However, in my replication, $\gamma_2$ is statistically significant on the 10% level with a point estimate of $0.022$.

The heteroskedasticity-robust standard errors reported in Table 3 are clustered at the level of the individual census block for Binder (2018), therefore allowing for correlated error terms across school years for each block. However, the standard errors reported for McMillen et al. (2017) are clustered at the level of the Safe Passage routes. I was not able to replicate this clustering, as it was unclear how exactly the clustering by routes was implemented for two reasons. First, a block can have multiple Safe Passage routes assigned to it. Second, the relevant route can change for a block between school years.[20]

Although I was able to replicate the main findings regarding the effect of the program on the crime counts, an obvious difference present in Table 3, is the number of observations used in the regression. Part of the discrepancy may be due to the fact that I use a newer version of the crime dataset. Blocks with constant (usually 0) crimes over the whole observation period are excluded from the regression due to the block fixed effects ($\delta_i$). In newer versions of the crime dataset, there might have been minor changes, as some crimes were additionally added or received GPS coordinates, which would make them available for the analysis. If, due to the increased number of crimes available, a block now experiences a non-constant crime count, it will no longer be excluded from the regression. However, this mechanism is unlikely to account for the full difference in observations. Therefore, there seems to be a difference in the creation of the estimation datasets, which I was not able to replicate. One other possibility would be the already mentioned restrictions on 5 school years prior and 3 school years after the implementation of the program. However, applied on my estimation dataset this would lead to significantly less observations than McMillen et al. (2017) have. It therefore seems unlikely to be the reason.[21]

---

[20]    For example, if a block is a "Three over block," but then a new Safe Passage route is established and crosses the "Three over block," it becomes a "Treated block," and its status is now defined by a different route.

[21]    The reduced estimation dataset yields 569,482 observations for the estimation of the effect on violent crime counts and 844,310 regarding property crime counts as the dependent variable.

# References

T. Ali. Is crime down on your kid's route to school? safe passage works, data shows. `https://www.dnainfo.com/chicago/20150908/bronzeville/crime-along-cps-safe-passage-routes-decreases-data-shows`, September 8 2015. Accessed: 23 April 2017.

A. A. Braga, A. V. Papachristos, and D. M. Hureau. The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice quarterly*, 31(4):633–663, 2014.

G. Broström. *glmmML: Generalized Linear Models with Clustering*, 2018. URL `https://CRAN.R-project.org/package=glmmML`. R package version 1.0.3.

C. da Costa-Luis, S. L., H. Mary, noamraph, M. Korobov, G. Chen, I. Ivanov, M. Bargull, M. D. Pagel, James, M. Nordlund, S. Malmgren, Socialery, J. McCracken, F. Dill, D. Panteleit, and Alexander. tqdm/tqdm: tqdm v4.23.4 stable, May 2018. URL `https://doi.org/10.5281/zenodo.1251290`.

R. Di Tella and E. Schargrodsky. Do police reduce crime? estimates using the allocation of police forces after a terrorist attack. *American Economic Review*, 94(1):115–133, 2004.

M. Draca, S. Machin, and R. Witt. Panic on the streets of london: Police, crime, and the july 2005 terror attacks. *American Economic Review*, 101(5):2157–81, 2011.

J. Grogger. The effects of civil gang injunctions on reported violent crime: Evidence from los angeles county. *The Journal of Law and Economics*, 45(1):69–90, 2002.

J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

J. Klick and A. Tabarrok. Using terror alert levels to estimate the effect of police on crime. *The Journal of Law and Economics*, 48(1):267–279, 2005.

T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, and C. Willing. Jupyter notebooks – a publishing format for reproducible computational workflows. In F. Loizides and B. Schmidt, editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87 – 90. IOS Press, 2016.

L. J. Krivo. Reducing crime through community investment: Can we make it work? *Criminology & Public Policy*, 13(2):189–192, 2014.

C. Lum and C. S. Koper. Evidence-based policing. In *Encyclopedia of criminology and criminal justice*, pages 1426–1437. Springer, 2014.

W. McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.

D. McMillen, I. Sarmiento-Barbieri, and R. Singh. Do more eyes on the street reduce crime? evidence from chicago's safe passage program. 2017. URL `https://ignaciomsarmiento.github.io/assets/Safe_Passage_WP.pdf`.

T. E. Oliphant. *A guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.

D. Weisburd, N. A. Morris, and E. R. Groff. Hot spots of juvenile crime: A longitudinal study of arrest incidents at street segments in seattle, washington. *Journal of Quantitative Criminology*, 25(4):443, 2009.

H. Wickham. *tidyverse: Easily Install and Load the 'Tidyverse'*, 2017. URL `https://CRAN.R-project.org/package=tidyverse`. R package version 1.2.1.