Supplementary Submission No 114a

INQUIRY INTO HIGH LEVEL OF FIRST NATIONS PEOPLE IN CUSTODY AND OVERSIGHT AND REVIEW OF DEATHS IN CUSTODY

Organisation: Public Interest Advocacy Centre

Date Received: 6 April 2021

---

12 March 2021

Adam Searle MLC
Chair, Select Committee into the High Level of First Nations People in Custody and Oversight and Review of Deaths in Custody
Parliament House
6 Macquarie St
Sydney NSW 200

Dear Select Committee,

Supplementary Submission – STMP Revised BOCSAR Data

At the public hearings before the Select Committee, evidence was taken about the findings in the BOCSAR Report entitled 'An evaluation of the Suspect Target Management Plan (STMP)' by Steve Yeong, first released 13 October 2020.

On 15 February 2021, BOCSAR published a revised version of this report. The revisions 'follow an internal review conducted by BOCSAR after three technical papers raised questions about aspects of the study methodology'[1].

I annex to this supplementary submission BOCSAR's 15 February revised evaluation, each of the critical reviews, and BOCSAR's response to the concerns raised in the critical reviews. All of these are available for download on the BOCSAR website.

Of particular relevance to the Select Committee's inquiry, the revised BOCSAR report finds significant increases in the risk of a custodial sentence post STMP: up 9.2 percentage points for the sample as a whole, up 10.0 percentage points for Aboriginal people and up 6.8 percentage points for juveniles.

The use of the STMP is demonstrated to increase the risk of imprisonment of the Aboriginal people it targets. If the Committee is serious about reducing the high rates of incarceration of Aboriginal people, it should call for the use of the STMP to be discontinued.

In submissions to the Select Committee, the following organisations have all raised concerns about the use of the STMP, or have called for the STMP to be discontinued for young people:

- The Aboriginal Legal Service NSW/ACT
- Legal Aid NSW
- Yfoundations
- Community Legal Centres NSW

[1] https://www.bocsar.nsw.gov.au/Pages/bocsar_publication/CJB233-Update-STMP.aspx

Leve 5, 175 L verpoo St
Sydney NSW 2000
Phone: 61 2 8898 6500
Fax: 61 2 8898 6555
www.p ac.asn.au
ABN: 77 002 773 524

We reiterate our call for the Committee to recommend the STMP be discontinued for young people.

Yours sincerely

Camilla Pandolfini
Principal Solicitor
Public Interest Advocacy Centre

Direct phone:
E-mail:

Encl:

1. An evaluation of the Suspect Target Management Plan, revised 15 February 2021.
2. Demonstrating an illusory reduction in crime via sampling, James Macdonald, 9 November 2020.
3. Simulation experiment for ''An evaluation of the Suspect Target Management Plan, October 2020 " Study, Dr. Gordana Popovic, 5 November 2020.
4. A critical review of the BOSCAR report: An evaluation of the Suspect Target Management Plan, Ian Watson, 9 November 2020.
5. Response to comments on 'An evaluation of the Suspect Target Management Plan - Crime Justice Bulletin 233', NSW Bureau of Crime Statistics and Research, February 2021.

Enclosed:

NSW Bureau of Crime Statistics and Research

This bulletin has been independently peer reviewed.

CRIME AND JUSTICE BULLETIN

NUMBER 233 (REVISED) | FEBRUARY 2021

An evaluation of the Suspect Target Management Plan

Steve Yeong

AIM

First, to determine whether STMP-II reduces violent and property crime. Second, to determine whether DV-STMP reduces Domestic Violence (DV) related crime. Third, to determine whether these programs operate through deterrence or incapacitation. Finally, to determine whether STMP-II has a differential impact on juveniles versus adults and Aboriginal Australians versus non-Aboriginal Australians.

METHOD

I use court outcome data in conjunction with the complete list of individuals subject to either program between May 2005 and September 2018. Using these data, I compare rates of offending and imprisonment in the 12 months before and after an individual becomes subject to either STMP program. In my analysis of STMP-II, I focus on specific violent and property offences to avoid reporting/detection bias from contaminating the estimates. In my analysis of DV-STMP, I also examine DV offences as they are the focus of the program. In order to determine whether either program operates through deterrence or incapacitation, I divide the sample into cohorts based on how much time each individual spent in custody during their follow up. Using these cohorts, I then estimate how much of the total crime reduction associated with either program can be explained by each cohort. I then apply these approaches to Aboriginal Australians and juveniles to explore whether these effects differ for both subgroups.

RESULTS

STMP-II is associated with large, practically and statistically significant reductions in property crime, but not violent crime. This crime reduction is also, however, accompanied by an increase in the risk of imprisonment. Although STMP-II is associated with increased imprisonment, any crime reduction benefit most likely occurs through deterrence rather than incapacitation. DV-STMP is associated with large, practically and statistically significant reductions in DV crime, but not other types of crime. DV-STMP has no association with the risk of imprisonment. When compared to adults, STMP-II is associated with larger crime reductions for juveniles, and a smaller, but significant, increase in the risk of imprisonment. When compared to non-Aboriginal people, STMP-II is associated with a smaller crime reduction benefit, and a larger increase in the risk of imprisonment for Aboriginal Australians. The associated crime reduction benefit for both groups is most likely due to deterrence.

CONCLUSION

Both STMP-II and DV-STMP are associated with a reduction in crime. In both cases, the associated crime reduction cannot be attributed to incapacitation. These claims also hold for juveniles and Aboriginal Australians.

KEYWORDS recidivism | incarceration | policing | focussed deterrence | deterrence | incapacitation

Suggested citation: Yeong, S. (2020). An evaluation of the Suspect Target Management Plan (Crime and Justice Bulletin No. 233 revised). Sydney: NSW Bureau of Crime Statistics and Research.

INTRODUCTION

Twenty-first century police employ a variety of tactics to maintain law and order. Arguably the most popular class of tactics employed by police are those rooted in focused deterrence. Focused deterrence, or "pulling levers" police programs typically involve the reallocation of existing police resources toward specific targets (e.g., physical spaces, individuals, gangs, types of crime) where police can get the greatest "bang for their buck". While there is a considerable body of research supporting the effectiveness of such tactics internationally, we know very little about whether and how such programs work in Australia.[1]

In New South Wales (NSW), the largest person-focused policing program is the Suspect Target Management Plan (STMP). STMP was introduced in 2002 with the objective of reducing crime by proactively policing individuals deemed to be at a high risk of offending. Following the release of a study by the NSW Youth Justice Coalition (Sentas & Pandolfini, 2017), STMP received considerable negative media attention.[2] Much of this criticism surrounding STMP centred on the targeting of vulnerable groups (i.e., juveniles and Aboriginal Australians). Despite this criticism, STMP continues to be one of the key elements of the NSW Police Force's strategy to reduce crime. The purpose of this study is to: determine whether STMP reduces crime; how STMP reduces crime; and finally, to examine if the effect of STMP is more pronounced for Aboriginal Australians and juveniles.

The Suspect Target Management Plan (STMP)

The Suspect Target Management Plan (STMP) is a NSW Police Force program that has been in operation since February 2002. The objective of STMP is to reduce crime by identifying individuals at a high risk of offending, notifying them that they are now subject to enhanced supervision and then proactively policing such individuals. In the context of STMP, proactive policing typically involves officers from the corresponding Police Area Command (PAC)[3] regularly conducting person and vehicle searches, bail compliance checks and issuing move-on directives.

There have been three iterations of STMP. STMP-I was introduced in February 2002, and then replaced by STMP-II in May 2005. The difference between STMP-I and STMP-II is the process an individual undergoes prior to being placed on STMP. Information regarding the selection process for STMP-I is unavailable. The selection mechanism for STMP-II is described shortly. DV-STMP is the third iteration of STMP. DV-STMP was introduced in October 2015 and sits alongside STMP-II. DV-STMP involves modifying elements of STMP-II in order to address the dynamics of Domestic Violence (DV).

STMP-II: May 2005 – present

The process begins when a member of the NSW Police Force nominates a Person of Interest (POI) for STMP. Any member of the NSW Police Force can nominate an individual for STMP and each PAC has its own unique STMP list. Once nominated, the corresponding PAC conducts a risk assessment. The risk assessment takes two factors into consideration: first, information regarding the POI's offending risk;[4] and second, the PAC's priority crimes.[5] Based on this assessment, a risk rating is generated for each candidate. Candidates can be of extreme, high, moderate or low risk. Following the risk assessment, candidates are then reviewed during the PAC's next Tasking and Deployment (T&D) meeting. It is at this meeting that the final decision regarding whether a candidate will go on STMP is made.[6] Candidates from all four-risk categories can be placed on STMP. The risk assessment is designed only to inform the discussion at each T&D meeting. Should a candidate be deemed suitable for STMP, an information report is generated

1 Braga, Weisburd and Turchan (2018) and Weisburd et al. (2019) provide an overview of this literature.
2 See for example Blanco (2018), McGowan (2017), O'Mallon (2017) or Shoebridge (2018).
3 A PAC is the name given to geographically defined police jurisdictions in NSW.
4 These factors include: prior offending; whether the use of violence and/or a weapon was involved in such offences; prior sentences of imprisonment; prior community based orders; whether the POI has addiction issues; whether the POI has known criminal associations; and finally, whether the POI was involved in crime from a young age.
5 Each PAC has its own list of priority crimes (e.g., domestic assault, break and enter), which can differ both between PACs and over time. I do not have information regarding each PAC's list of priority crimes.
6 I am not able to observe information relating to cases where an individual is considered for STMP at a T&D meeting and deemed unsuitable.

for the “target”. This information report is then allocated to a team of officers within the PAC. The team leader is responsible for designing a Target Action Plan (TAP) to deal with the target.[7] Targets currently on STMP are reviewed at each T&D meeting. Targets are removed from an STMP list if they have died, been incarcerated[8] or if their criminal behaviour appears to have ceased.[9]

DV-STMP: October 2015 – present

DV-STMP involves four modifications to STMP-II. First, if an individual has been identified in multiple DV incidents, police are encouraged to nominate the POI for DV-STMP. Second, additional information is considered during a T&D meeting for DV-STMP candidates (e.g., whether children are involved). Third, the TAP involves contacting the victim to inform them that the corresponding POI has been placed on STMP. And finally, police also contact the local police prosecutor(s) and inform them that the POI has been placed on STMP.

Literature

A question of first order importance to policymakers is whether police reduce crime. More than two decades of empirical research indicate that the answer to this question is yes. An increase in police numbers generates a reduction in crime.[10] Although the magnitude and precision of estimates differ between jurisdictions, Chalfin and McCrary (2017) suggest that, in general, a one per cent increase in police numbers generates a 0.4 and 0.2 per cent reduction in violent and property crime, respectively. A question of second order importance to policymakers is how police reduce crime. There are two channels through which police may reduce crime (Becker, 1968). The first is referred to as incapacitation: the crime reduction that occurs when offenders are unable to offend due to their imprisonment.[11] The second is referred to as deterrence, of which there are two types. General deterrence refers to the idea that police reduce crime by lowering an individual’s proclivity to offend through fear of apprehension and punishment. Specific deterrence refers to the idea that individuals who are arrested, sanctioned or subjected to supervision will be deterred from further offending. Focused deterrence is one of the various measures that criminologists consider to fall within the specific deterrence category.

Focused deterrence refers to the relocation of police resources toward a relatively small number of offenders responsible for a disproportionately large fraction of crime. Focused deterrence programs typically involve three elements. The first of which is to communicate an explicit message of deterrence to those targeted by the intervention. The purpose of the message is to convey the idea that the certainty, severity and swiftness of apprehension and sanctions have now increased. This message is often conveyed by members of the local community working in conjunction with law enforcement. The second and third elements refer to intensive police supervision and the increased availability of social services (e.g., social housing, job training and education programs).

Arguably the most famous focused deterrence policing deployment was the 1990s Boston Operation Ceasefire (BOC). Like many large U.S. cities, Boston experienced a sharp increase in youth homicide in the late ‘80s and early ‘90s. A working group of police, youth workers and academics determined that one per cent of the city’s youth (many of whom were active gang members) were responsible for more than 60 per cent of youth homicides. The BOC was designed as a response to this problem. The BOC involved police, youth workers, churches, probation and parole officers working together to communicate an explicit message to gang members that violent crime would not be tolerated. All parties involved made it clear that participation in violent crime would generate an immediate and aggressive response from police.

7 A TAP typically involves notifying the target that they are now subject to enhanced supervision and making regular contact through move along directives and person and vehicle searches. I do not have information regarding specific TAPs.
8 In practice, police often do not take individuals off STMP after they are imprisoned. This is one of the reasons I limit the analysis to the first 12 months an individual becomes subject to STMP.
9 If a target moves to the jurisdiction of another PAC, the case is transferred to the new PAC.
10 Studies within this field of research typically leverage exogenous city or state level variation in police numbers to identify this causal relationship. Researchers have applied Instrumental Variables strategies (Evans & Owens 2007; Levitt, 1997; Owens, 2013; Yeong, 2019), Difference-in-Differences (Di Tella & Schargrodsky, 2004; Machin & Marie, 2011), high frequency time series (Chalfin & McCrary, 2018; Corman & Mocan, 2000; Marvell & Moody, 1996) and other approaches (DeAngelo & Hansen, 2014; Klick & Tabarrok, 2005; Shi, 2009), and consistently found that increasing the number of police reduces crime.
11 To the best of my knowledge there are only two quasi-experimental studies examining how an increase in aggregate police numbers generates a reduction in crime (Owens, 2013; Yeong, 2019), both of which support the idea that police reduce crime through deterrence rather than incapacitation.

Participation in non-violent crime, however, would entail a "business-as-usual" response. Simultaneously, these groups also offered increased levels of social services (e.g., employment training and education programs) to gang members. While the magnitude of the BOC's effect has been the subject of much debate (Braga et al., 2001; Braga et al., 2014; Rosenfeld et al., 2005), the general consensus is that the BOC generated a sizable, significant reduction in youth homicides.

The success of BOC set the tone for a variety of focused deterrence interventions. Braga et al. (2018) divide such interventions into three groups.[12] The first two groups focus on addressing gang-related violent crime and drug market activity. Like BOC, these interventions typically involve expressing a clear threat of punishment to selected high-risk offenders, while simultaneously working to increase access to social services. Messages of deterrence and access to social services are communicated through members of the community working in conjunction with law enforcement. Evaluating such programs presents two challenges, the first of which is reporting bias and the second is selection bias.

Reporting bias refers to crime that goes unreported to, or detected by, police. To address this issue, criminologists investigating the effect of gang related violent crime interventions limit their analysis to serious violent crime (e.g., homicide, assault resulting in serious injury, shootings and stabbings) that rarely go undetected by police. Unfortunately, addressing issues regarding detection bias for interventions targeting drug market operations is less straightforward as drug offences (e.g., use/possess/supply drugs) are heavily influenced by the level of police activity.

Selection bias, in the current context, refers to the idea that gangs become subject to an intervention precisely because of an ex-ante crime problem. This makes finding a valid counterfactual difficult because targeted gangs are, by construction, at a higher offending risk than non-targeted gangs. Much of the early research examining violent gang interventions (Boyle et al., 2010; Braga, 2008; Braga et al., 2008; Engel et al., 2013; McGarrell et al., 2006) and drug market interventions (Corsaro et al., 2010a) relied principally on time series variation.[13] Criminologists have, however, quickly adopted more sophisticated panel data approaches to examine the effect of both violent crime interventions (Braga et al., 2013; Circo et al., 2020; Grunwald & Papachristos, 2017; Papachristos et al., 2007) and drug market interventions (Corsaro et al., 2012; Corsaro et al., 2013; Saunders et al., 2015). Such approaches address the selection bias problem by taking pre-existing differences between treatment and control groups into account, and then comparing offending rates, net of this adjustment, before and after the intervention. The consensus in this literature is that both types of interventions are effective in generating moderate reductions in crime (Weisburd et al., 2019).

There is, however, very little rigorous empirical evidence with regard to the third type of focused deterrence intervention: those targeting specific individuals. This is likely due to the lack of available (individual level) micro data. In fact, to the best of my knowledge, the Community Initiative to Reduce Violence (CIRV) in Glasgow is the only individual focused deterrence program to have been rigorously evaluated.

The CIRV was introduced in 2002 to reduce knife related crime. Under the CIRV, offenders residing in Glasgow's "east end" were identified by police using operational intelligence. Police targeted young, male gang members with a history of violent offending. These individuals were then provided with a phone number they, or any member of their gang, could use to contact a "street worker". Following contact, the street worker would obtain a written commitment from the individual to abstain from violence and not carry a weapon. The incentive to participate lay in the provision of additional social services offered through the program (e.g., employment training, education programs, public housing). Detected breaches

12 Weisburd et al. (2019) divide focused deterrence programs into four groups, based on their operational characteristics. Hot spot policing refers to interventions that direct police resources toward geographical areas abundant with crime (e.g., a particular neighborhood or street). Problem-solving strategies involve augmenting existing policy settings to address factors associated with particular types of crime (e.g., closing bars early to avoid alcohol related violence). Community-based strategies centre around involving the local community in the crime prevention effort (e.g., through the use of neighborhood watch). And finally, offender-focused policing refers to the reallocation of police resources toward individuals responsible for a disproportionately large fraction of crime.

13 Although these studies also use time series approaches to look for a structural break in areas not subject to an intervention, they do not concatenate such information into a single Difference-in-Differences style model.

of their written commitment would result in an immediate freeze on any service they were receiving. To examine the effectiveness of the CIRV, Williams et al. (2014) compared the behaviour of individuals participating in the CIRV to a matched control group of individuals residing in Glasgow's "south side". These individuals were matched using the same criteria the police used to identify individuals subject to the CIRV. Williams et al. (2014) then compared the offending of individuals within each group in the 12 months immediately before and after the program. Williams et al. (2014) found that the CIRV generated significant reductions in both violent crime and the probability of carrying a knife.

In contrast to the CIRV, and the vast majority of programs referenced in this section, neither STMP-II nor DV-STMP involve increasing access to social services. Similarly, neither program involves members of the local community working with police to communicate an explicit message of deterrence. These departures from the conventional focused deterrence approach beg the question of whether STMP is likely to be as effective as the programs described in this section. The purpose of this study is to shed light on this important question.

THE CURRENT STUDY

The current study is concerned with answering the following three research questions: (1) Is STMP-II or DV-STMP associated with a reduction in crime? (2) Is the crime reduction benefit associated with each program achieved through deterrence or incapacitation? (3) Do the answers to questions (1) and (2) differ for juveniles or Aboriginal Australians?

Data

I utilise two datasets in this study. The first is an extract from the NSW Bureau of Crime Statistics and Research's Reoffending Database (ROD). The ROD extract contains information relating to all legal proceedings finalised in a NSW criminal court over the period 1 January 1996 to 30 September 2019.[14] For each individual in ROD, I am able to observe their date of birth (and death, if they died); gender; MSPDI;[15] CNI;[16] Aboriginality;[17] a complete history of custodial episodes (i.e., I can observe dates that they entered and exited custody for both remand and sentences of imprisonment); and finally, information relating to any finalised court appearance. For each finalised court appearance, I am able to observe the nature of each charge (e.g., assault, motor vehicle theft;[18] the PAC responsible for charging the POI; the date that the offence was alleged to have taken place; and the outcome of the charge (i.e., guilty or not guilty)).

The second dataset is an extract from the NSW Police Force's Computerised Operational Policing System (COPS). The COPS extract contains information relating to all individuals placed on either STMP-II or DV-STMP between 1 May 2005 and 1 September 2019. For each individual placed on STMP, I am able to observe their CNI; date of birth; PAC responsible for placing them on STMP as well as the date they were placed on and taken off STMP.[19]

Individuals are identified in the police dataset by CNI. Individuals can be identified in ROD using either the CNI or MSPDI. I merge the two datasets using the CNI. In some cases, however, there are multiple CNIs for

14 Criminal proceedings begin when a member of the NSW Police Force charges an individual with one or more offences. After this occurs, the charge(s) must be finalised in a criminal court. If the defendant is found "not guilty", then finalisation occurs on the day this determination is made. If the defendant is found "guilty", then finalisation occurs on the day that the sentence (i.e., the penalty) is handed down.

15 The MSPDI is a randomized numerical code used to identity individuals with a finalised court appearance in ROD.

16 The CNI or Central Names Index is an individual level identifier given to individuals associated with an alleged offence.

17 That is, a binary variable equal to one if the individual has ever identified as Aboriginal to police, zero otherwise.

18 With one exception, the nature of each charge is determined using the Australian and New Zealand Offence Classification (ANZSOC) codes. Interested readers are directed to ABS (2011) for further information regarding ANZSOC codes. The exception is DV related charges. DV related charges are identified using the law part code associated with each charge. Law part codes are used in NSW to identify specific types of charges. Interested readers are directed to the Judicial Commission of NSW (2020) for more information pertaining to law part codes.

19 Some individuals are placed on STMP for an unreasonably long period of time (14 years in some instances). This is likely to be due to the police failing to take such individuals off STMP after their imprisonment. For this reason, I limit my analysis of offending to the year immediately before and after an individual is placed on STMP.

a single MSPDI. This is likely due to inconsistent recording of individuals. For this reason, after merging the two datasets, I use the MSPDI to identify unique individuals.

In total, there are 12,059 unique individuals (i.e., MSPDIs) in the COPS dataset, 1,534 of which have been placed on DV-STMP, 10,667 that have been placed on STMP-II and 142 that have been subject to both programs. Of the total 12,059 individuals, I was unable to match 45 individuals between datasets. This could be due to either incorrect information in at least one of the datasets or because the individual has never been formally charged by police. In four instances, the PAC responsible for placing an individual on STMP is also not recorded in COPS. I drop these observations from the sample completely. I then further restrict the sample to individuals who were either: placed on STMP-II between 1 May 2005 and 30 September 2018; and/or placed on DV-STMP between 30 October 2015 and 30 September 2018. This allows me to observe the offending behaviour of each individual on STMP for at least 12 months and results in an attrition of 932 individuals. The net result is a dataset that contains information for 10,106 individuals subject to STMP-II, 1,028 individuals subject to DV-STMP and 56 individuals who were subject to both programs.

Descriptive statistics

Table 1 provides descriptive information for individuals placed on STMP-II (in Panel A) and DV-STMP (in Panel B) at two points in time: 365 days prior to being placed on STMP, and the first (free)[20] day the individual was subject to STMP. From Table 1 we can see that individuals subject to STMP-II (DV-STMP) are, on average, about 18 (22) years old at first contact with the criminal justice system (CJS), predominately male and (disproportionately) Aboriginal in 45.8 (37.1) per cent of cases. By the time that the typical individual is placed on either form of STMP, he has almost 10 prior court appearances, half of which relate to the use of violence, one relating to the use of weapons and two relating to the use of drugs. He has also had a sentence of imprisonment[21] and four community orders, all by age 26 (35 for DV-STMP).[22] It is also worth mentioning that 26.9 (2.87) per cent of individuals placed on STMP-II (DV-STMP) were below the age of 18 when placed on the program.

20 That is, the first day after being placed on STMP that the individual was not in custody and therefore able to offend in the community. 422 individuals started STMP-II during a custodial episode shortly before being released into the community.

21 Juvenile control orders are also counted as a sentence of imprisonment. Note that in a previous version of CJB233 juvenile control orders were incorrectly defined as a community based penalty. This error has now been corrected and all tables throughout this bulletin have been updated accordingly. The primary implication of this error was that the previous version of this bulletin stated that STMP-II was not associated with an increased risk of imprisonment for juveniles.

22 Counts of prior offences refer to both proven and unproven offences (i.e., the number of offences with which the police have charged the individual, regardless of whether the individual was found guilty).

Table 1. Individuals placed on STMP between May 2005 and September 2018

	One year before STMP			First day on STMP			Difference	
	N	Mean	Std. Dev.	N	Mean	Std. Dev.	Estimate	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. STMP-II								
Age	10,100	24.944	10.219	10,100	25.961	10.223	1.017***	(0.144)
Age at first CJS contact	10,080	17.941	7.001	10,080	17.941	7.001	-	-
Male	10,103	0.915	0 279	10,103	0.915	0.279	-	-
Aboriginal	10,085	0.458	0.498	10,085	0.458	0.498	-	-
Prior court appearances	10,103	8.210	7.626	10,103	9.703	7.639	1.493***	(0.107)
Prior violent offences	10,100	3.617	4.878	10,100	4.500	5.157	0.883***	(0.071)
Prior weapon offences	10,100	0.396	1.161	10,100	0.544	1.383	0.148***	(0.018)
Prior drug offences	10,100	1.309	2.440	10,100	1.661	2.768	0.352***	(0.037)
Prior community orders	10,103	4.704	3.102	10,103	4.704	3.102	-0.000	(0.044)
Prior YJCs and cautions	10,103	0.526	1.055	10,103	0.692	1.180	0.166***	(0.016)
Prior prison sentences	10,103	1.741	3.087	10,103	1.995	3.265	0.254***	(0.045)
Panel B. DV-STMP								
Age	1,028	34.031	10.392	1,028	35.048	10.385	1.017*	(0.458)
Age at first CJS contact	1,027	22.003	9 344	1,027	22.003	9.344	-	-
Male	1,028	0.914	0 280	1,028	0.914	0.280	-	-
Aboriginal	1,026	0.371	0.483	1,026	0.371	0.483	-	-
Prior court appearances	1,028	9.529	8 372	1,028	10.924	8.540	1.395***	(0.373)
Prior violent offences	1,028	6.675	7.469	1,028	8.894	7.621	2.219***	(0.333)
Prior weapon offences	1,028	0.408	1.162	1,028	0.505	1.297	0.097	(0.054)
Prior drug offences	1,028	1.278	2.149	1,028	1.523	2.379	0 245*	(0.100)
Prior DV offences	1,028	4.371	6.162	1,028	8.129	6.951	3.759	(0.290)
Prior community orders	1,028	4.340	3.044	1,028	4.340	3.044	0.000	(0.134)
Prior YJCs and cautions	1,028	0.403	0.979	1,028	0.420	0.998	0.018	(0.044)
Prior prison sentences	1,028	1.589	3.054	1,028	1.871	3.220	0 281*	(0.138)

Note. N = Observations, DV = Domestic Violence, YJC = youth justice conference, CJS = criminal justice system, robust standard errors in parentheses, **** *p*<0.001, ** *p*<0.01, * *p*<0.05.

From Table 1 we can also see significant increases in the number of court appearances (involving violence, drugs or weapons), prison sentences and community orders in the year leading up to being placed on STMP. While Table 1 suggests that the police are identifying high-risk individuals for STMP, it does make finding a valid counterfactual difficult. That is, finding a suitable control group for individuals placed on STMP is difficult because an individual's offending risk increases so sharply in the year before being placed on STMP. This is discussed at length in the proceeding section.

Empirical approach

Estimating the causal relationship between STMP and crime is confounded by two factors.

1. **Detection bias:** Holding the actual level of offending constant, once an individual is placed on STMP they are more likely to be caught offending.
2. **Selection bias:** When compared with individuals not on STMP, individuals on STMP are likely to be at a higher risk of offending, irrespective of whether STMP has any impact on offending.

In order to deal with the detection bias problem, I limit the analysis to specific types of violent and property crime least likely to be influenced by policing or surveillance.[23] These violent crimes include homicide, assault occasioning grievous bodily harm and robbery. The property crimes include theft, motor vehicle theft and break and enter. Although DV offences may be influenced by detection bias, I also investigate these offences as they are the focus of DV-STMP. In order to deal with the selection bias problem, I limit the estimation sample to individuals (who will eventually be) subject to STMP, and then employ the approach described below.[24]

Without loss of generality, suppose that we are interested in estimating the effect of STMP on the probability of at least one proven offence within a POI's first 12 months of STMP. I begin by restricting the estimation sample to individuals who will (at some point) be subject to STMP. Then, for each individual, I further restrict the estimation sample to the year immediately before and after they begin STMP (i.e., two years per individual). Finally, I then compare the offending behaviour of (treated) individuals subject to STMP in period t, with (control) individuals who will be subject to STMP in period $t+1$.

This approach, which can be generalised to other outcome measures used in this study (e.g., the probability of being incarcerated or the number of offences within 12 months) is summarised in Equation 1.

$$y_{ipt} = \beta_0 + \beta_1 Post_{ipt} + \gamma X'_{it} + \lambda_{pt} + u_{ipt} \qquad (1)$$

In Equation 1, i indexes an individual, p indexes the PAC responsible for placing the individual on STMP and t indexes a month-year combination. y_{ipt} is a binary variable equal to one if an individual offends within 12 months of STMP, zero otherwise. $Post_{ipt}$ is a binary variable equal to one for periods after an individual becomes subject to STMP, zero otherwise. Importantly, individuals become subject to STMP at different points in time. Therefore, $Post_{ipt}$ varies both within and between individuals subject to STMP. This is discussed in more detail shortly. The set of control variables, represented by X'_{it}, includes the individual characteristics from Table 1. These control variables directly relate to the selection criteria for STMP outlined earlier. Recall, however, that selection for STMP is also a function of each PAC's priority crimes (which are unobservable). For this reason, I also include a set of PAC-by-year fixed effects denoted by λ_{pt}. These fixed effects render the estimates robust to PAC-specific considerations such as their priority crimes, annual budgeting allocations, variation in the application of STMP, local labour market conditions and the demographic characteristics of civilians living within the jurisdiction of each PAC. The error term is represented by u_{ipt}, and all other terms are coefficients to be estimated.

The coefficient of interest, β_1, is identified through variation in the timing of when individuals become subject to STMP. In order for β_1 to have a causal interpretation, an individual's risk of offending must be conditionally unrelated to this timing.[25] Given that the timing of when an individual becomes subject to STMP is a direct function of their offending behaviour, there is no reason we should expect this condition to hold.

This problem is illustrated in Figures 1a and 1b. Figure 1a plots the daily probability of at least one selected (proven) violent or property crime for individuals subject to STMP-II, and Figure 1b plots the daily probability of at least one (proven) DV offence. Both figures report this information over the 365 calendar days before and after an individual is placed on STMP.[26] If β_1 were to have a causal interpretation, we would expect so see no trend in offending prior to STMP, followed by a sharp (downward) trend after placement on STMP.

23 These crimes are least likely to be influenced by reporting and detection biases for two reasons. First, they are not discovered or experienced by police (like drug possession and offensive behavior for example) and second, victims have a clear incentive to report such crimes to police. Limiting the analysis to these offences is also consistent with prior research, described in the literature review.

24 The fact that individuals exert no influence over whether they are placed on STMP (other than through their offending behaviour) lends itself to a matching strategy. This idea is thoroughly investigated in the Appendix. Interestingly, I was not able to find a credible match for individuals subject to STMP using the entire Reoffending Database (which contains information for every person charged by the NSW Police Force since 1996). This suggests that the people the police select for STMP are truly distinct from other offenders they interact with.

25 Interested readers are directed to Cobb-Clark et al. (2018) and Hoynes and Schanzenbach (2009) for an introduction to identification in a "rolling" Difference-in-Differences setup.

26 Figure A2 in the Appendix reports this information over the 730 days before and after the individual is subject to STMP. The general pattern observed in Figures 1a and 1b remains unchanged.

However, from Figures 1a and 1b we can see sizable upward trends in the year leading up to STMP, followed by sharp downward trends immediately after being placed on STMP. This finding has two implications for the analysis. First, my estimates do not have a causal interpretation. Instead, they must be interpreted as the association between STMP and offending. And second, this would suggest that the police are both correctly identifying individuals at a high risk of offending for STMP, and that once placed on STMP, an individual's risk of offending drops dramatically.

The fact that an individual's offending risk is increasing in the lead up to STMP suggests that β_1 may actually be *underestimating* the true casual effect of STMP on crime. This is contingent upon the assumption that an individual's offending risk would have not decreased in the absence of STMP. One could argue that this is not likely to hold because standard criminal justice responses (e.g., imprisonment or other sanctions) would have generated a reduction in crime absent STMP. However, I do not believe this to be the case since these standard responses were in effect in the year prior to an individual's placement on STMP. From Table 1 we know that in the year leading up to STMP, individuals were indeed subject to imprisonment and other sanctions while their offending continued to rise nonetheless. We can also rule out systematic (state-wide) changes to the CJS as individuals in our sample are placed on STMP at different points in time. As such, the estimates reported in this study are likely to be conservative. That is, they underestimate the true impact of STMP's crime reduction benefit.

Figure 1a. Daily rates of violent and property crime for STMP-II

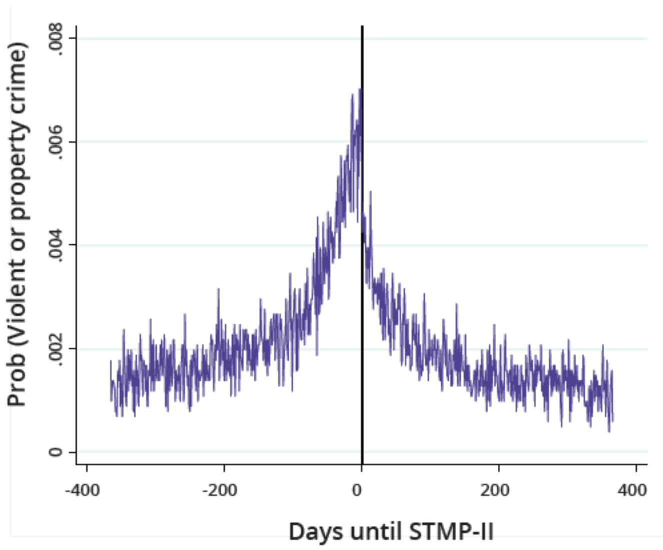
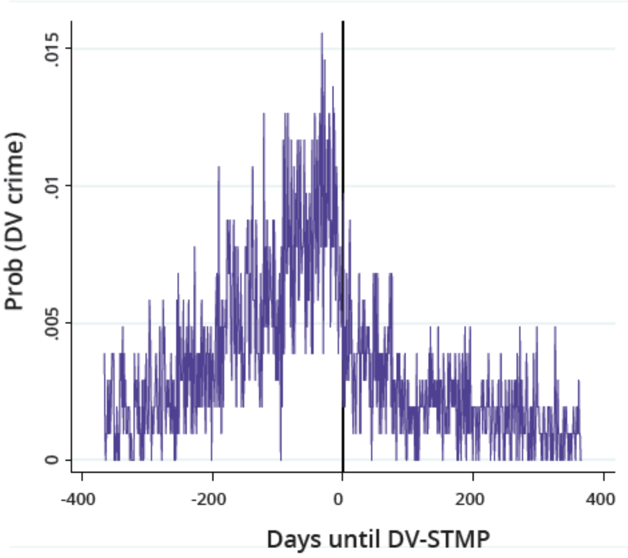


Figure 1b. Daily rates of DV crime for DV-STMP



If STMP is generating a reduction in crime, the question is how? In order to work toward answering this question, Figures 2a and 2b plot the daily probability of a custodial episode (remand or sentenced) in the 365 calendar days before and after an individual becomes subject to STMP. From Figure 2a we can see that about 20 per cent of individuals subject to STMP-II experience a custodial episode in the lead up to STMP-II, with this number falling as individuals begin STMP-II. After placement on STMP-II, we can see a sharp increase to almost 25 per cent.[27] This suggests that many individuals are placed on STMP-II following a custodial episode before returning to custody after placement on STMP-II. Taken together, this indicates that at least some of the crime reduction benefit observed in Figure 1a is driven by incapacitation.

27 Approximately three per cent of individuals are placed on STMP during a custodial episode shortly before being released into the community. For some individuals this occurs following an arrest while they are held on remand, and for others this occurs shortly before being released from a prison sentence. In either case, I only begin tracking their offending behaviour (in the regressions) after they have been released into the community.

From Figure 2b we can see a similar, although more volatile, pattern prior to placement on DV-STMP. Following placement on DV-STMP, however, we can see a lower rate of custodial episodes. When interpreted in conjunction with Figure 1b, this indicates that DV-STMP is working predominately through deterrence. In the next section, I report estimates that quantify this graphical analysis.

Figure 2a. Daily rates of time spent in custody for STMP-II

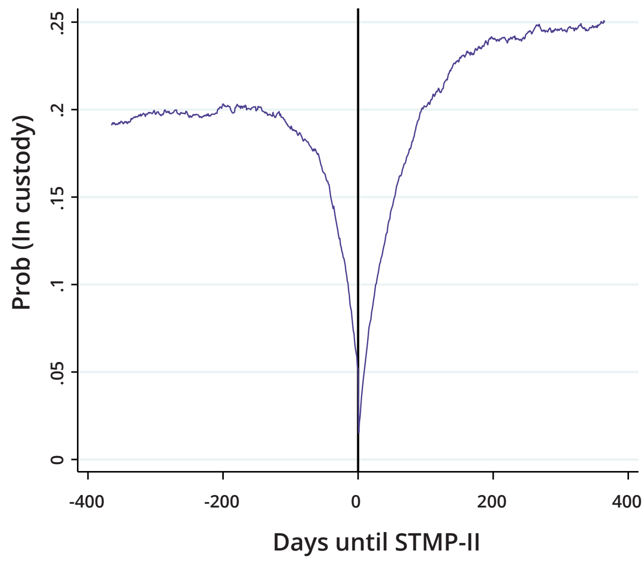
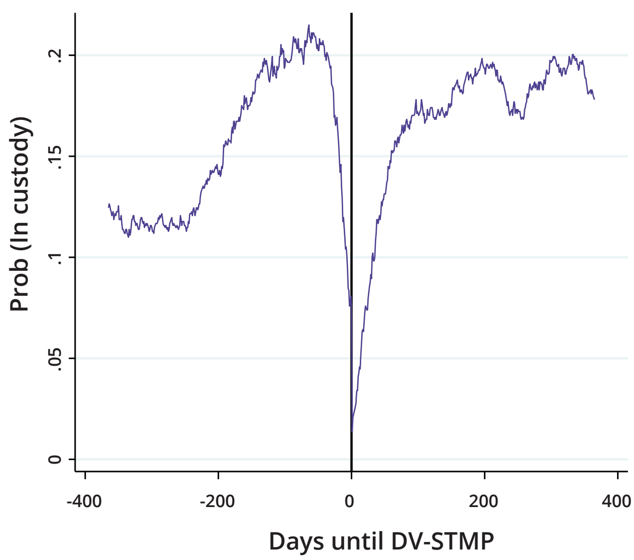


Figure 2b. Daily rates of time spent in custody for DV-STMP



RESULTS

STMP-II

Table 2 reports Ordinarily Least Squares (OLS) estimates with robust standard errors clustered at the PAC-by-year level for STMP-II.[28] Columns 1, 2 and 3 report estimates from an OLS regression of Equation 1 that includes: no controls or fixed effects; control variables; control variables and PAC-by-year fixed effects, respectively. Columns 4 and 5 then limit the estimation sample to individuals who were below the age of 18 when placed on STMP-II, and have identified as Aboriginal, respectively.

From Panel A we can see that STMP is associated with a large, statistically and practically significant reduction in the likelihood of at least one violent or property offence within 12 months. Moving from columns 1 to 3, we can see that inclusion of the control variables and fixed effects decreases the magnitude of the association by about 26.5 per cent. The complete model (in column 3) suggests that STMP-II is associated with a 6.1 percentage point (pp) reduction in the probability of at least one violent or property crime within 12 months. In relative terms, expressed as a fraction of the rate at which offenders one year away from STMP-II offend, this equates to a decrease of about 14.6 per cent.[29] From columns 4 and 5, we can see that the effect of STMP-II is heterogeneous. The association is more pronounced for juveniles (a reduction of 14.2 pps or 21.7%) and is less pronounced for Aboriginal people (a reduction of 4.3 pps or 9.2%).[30]

28 I employ an OLS regression instead of a Probit or Logit regression because imprisonment is an extremely rare event for juveniles. As such, in a regression with PAC fixed effects, the 75 per cent of PACs that have never imprisoned a juvenile are dropped from the regression. In Table A6 of the Appendix I report average marginal effects from a Probit regression of Equation 1 for the full sample of observations (i.e., a Probit regression analogous to column 3 in Tables 2 and 3). This generates no meaningful change to the main results. This is not surprising given that OLS estimators have an interpretation robust to non-linearities induced by binary dependent variables. As outlined by Angrist and Krueger (2001), the main advantages presented by competing non-linear models are incurred when the objective is prediction not causal inference.

29 The pre-STMP rate of offending is given by the row labeled "control group mean" in Tables 2 and 3. A complete set of descriptive statistics for all outcome variables examined in this study is available for interested readers in Table A4 of the Appendix.

30 Figure A4 in the Appendix reports figures analogous to Figures 1 and 2 for these subgroups.

Panels B and C examine violent and property offences, respectively. From Panel B we can see that prior to placement on STMP-II, only five per cent of individuals committed a selected violent crime in the 12 months prior to STMP-II. This is likely due to the severity of such crimes (i.e., homicide, assault occasioning grievous bodily harm and robbery). We can also see that there is no significant relationship between STMP-II and offending for these crimes when averaged across all participants. This is not the case for juveniles in our sample. Prior to placement on STMP-II, 13 per cent of juveniles had committed at least one of these violent crimes. After placement on STMP-II, this decreases by 2.9 percentage points (21.8% in relative terms).

The estimates reported in Panel C are largely consistent with their counterparts in Panel A. This suggests that the bulk of the reduction in crime associated with STMP is driven by a reduction in property crime.

Table 2. The relationship between STMP-II and the probability of offending and imprisonment

	(1) Naive	(2) Controls	(3) PAC fixed effects	(4) Juveniles	(5) Aboriginal
Panel A.	-0.083***	-0.061***	-0.061***	-0.142***	-0.043***
Selected violent or property crime	(0.006)	(0.006)	(0.006)	(0.015)	(0.010)
Control group mean	0.418	0.418	0.418	0.654	0.468
Observations	20,200	20,120	20,120	5,460	9,192
Adjusted R-squared	0.007	0.010	0.010	0.063	0.093
Panel B.	-0.004	0.001	0.001	-0.029**	0.009
Selected violent crime	(0.003)	(0.003)	(0.003)	(0.010)	(0.005)
Control group mean	0.050	0.050	0.050	0.133	0.052
Observations	20,200	20,120	20,120	5,460	9,192
Adjusted R-squared	0.000	0.032	0.031	0.003	0.021
Panel C.	-0.082***	-0.062***	-0.062***	-0.121***	-0.046***
Selected property crime	(0.006)	(0.006)	(0.006)	(0.014)	(0.009)
Control group mean	0 393	0.393	0.393	0.594	0.445
Observations	20,200	20,120	20,120	5,460	9,192
Adjusted R-squared	0.007	0.090	0.090	0.07	0.086
Panel D.	0.115***	0.092***	0.092***	0.068***	0.100***
Imprisonment	(0.006)	(0.006)	(0.006)	(0.011)	(0.009)
Control group mean	0.189	0.189	0.189	0.131	0.224
Observations	20,206	20,120	20,120	5,460	9,192
Adjusted R-squared	0.018	0.091	0.090	0.151	0.093
Controls	N	Y	Y	Y	Y
PAC-by-year fixed effects	N	N	Y	Y	Y

Note. PAC = Police Area Command, robust standard errors clustered at the PAC-by-year level in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Finally, Panel D examines the likelihood of a prison sentence within 12 months of STMP-II.[31] From column 3 we can see STMP-II is associated with a 9.2 percentage point increase in the likelihood of a prison sentence. In relative terms, this increases the probability of a prison sentence by 48.7 per cent. This association is, in absolute terms, slightly larger for Aboriginal people on STMP-II (an increase of 10 pps or 44.6%) and slightly weaker for juveniles on STMP-II (an increase of 6.8 pps or 51.9%).

DV-STMP

I examine the relationship between DV-STMP and crime and imprisonment in Table 3. Table 3 examines the probability of a DV crime in Panel A, a property or violent crime in Panel B, a violent crime in Panel C, a property crime in Panel D, and finally imprisonment in Panel E. I do not report estimates among the subset of juveniles or Aboriginal Australians on DV-STMP as there are too few observations to draw any reasonable conclusions. From Panel A we can see that DV-STMP is associated with a 29.7 percentage point reduction in the probability of a DV crime within 12 months (41.2% in relative terms). Unfortunately, I am unable to determine whether this reduction is the result of DV-STMP or reporting bias. Given that the offender is explicitly notified that he has been placed on STMP, it is reasonable to assume that he may now more closely monitor the reporting behaviour of the victim. As such, these results should be interpreted with caution.

In either case, the fact that DV-STMP generates such a large reduction in DV crime begs the question of where this reduction is coming from. From Panels B, C and D we know that the reduction is not being generated through the selected violent and property crimes outlined earlier. To answer this question, in Table A3 of the Appendix, I divide DV offences into several categories. These categories include various definitions of assault, sexual offences, property crime and offences against justice procedures (e.g., breaching an Apprehended DV Order). Worth mentioning is that these categories include all types of DV offences, including those that may be subject to reporting/detection bias. This analysis suggests that the reduction in DV crime is driven by reductions in assault, property crime and breaches of court orders.

The final outcome examined in Panel E of Table 3 is the probability of imprisonment. Here we can see that, net of controls and fixed effects, DV-STMP has no significant relation to the probability of being incarcerated.

31 It is worth pointing out that a prison sentence, which I refer to as “imprisonment”, differs from a custodial episode. A custodial episode also includes time spent in custody when an individual is refused bail and therefore held on remand.

Table 3. The relationship between DV-STMP and the probability offending and imprisonment

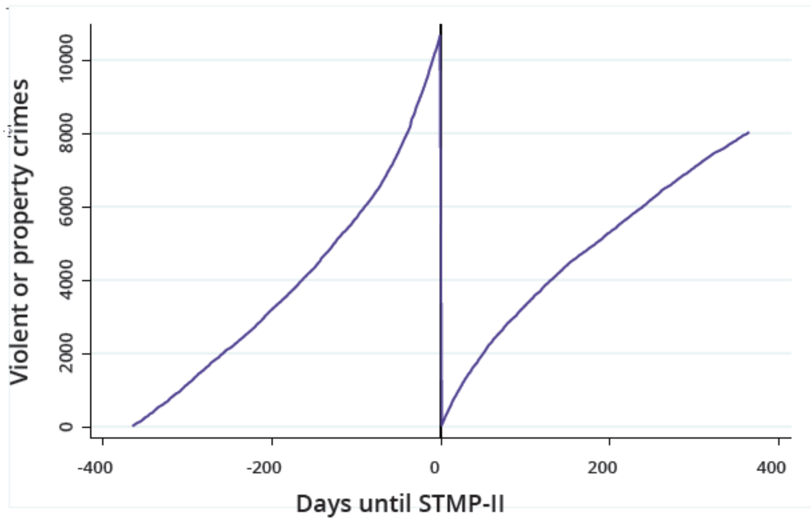
	(1) Naive	(2) Controls	(3) PAC fixed effects
Panel A.	-0.303***	-0.296***	-0.297***
DV crime	(0.019)	(0.021)	(0.021)
Control group mean	0.721	0.721	0.721
Observations	2,056	2,050	2,050
Adjusted R-squared	0.093	0.129	0.128
Panel B.	-0.013	-0.000	-0.002
Selected violent or property crime	(0.012)	(0.012)	(0.012)
Control group mean	0.091	0.091	0.091
Observations	2,056	2,050	2,050
Adjusted R-squared	0.001	0.045	0.043
Panel C.	-0.002	-0.001	-0.001
Selected violent crime	(0.003)	(0.003)	(0.003)
Control group mean	0.007	0.007	0.007
Observations	2,056	2,050	2,050
Adjusted R-squared	0.000	0.011	0.008
Panel D.	-0.012	-0.000	-0.002
Selected property crime	(0.011)	(0.011)	(0.011)
Control group mean	0.087	0.087	0.087
Observations	2,056	2,050	2,050
Adjusted R-squared	0.001	0.044	0.042
Panel E.	0.045*	0.002	0.001
Imprisonment	(0.019)	(0.020)	(0.020)
Control group mean	0.220	0.220	0 220
Observations	2,056	2,050	2,050
Adjusted R-squared	0.003	0.088	0.087
Controls	N	Y	Y
PAC-by-year fixed effects	N	N	Y

Note. PAC = Police Area Command, robust standard errors clustered at the PAC-by-year level in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Deterrence vs. incapacitation

From Table 2 we know that STMP-II is associated with both a lower probability of offending and an increased probability of incarceration. The question I address in this section is whether the association between STMP-II and crime is driven by deterrence or incapacitation. To answer this question, instead of examining the relationship between STMP-II and the *probability* of at least one selected violent or property crime, I focus on the *number* of selected violent and property crimes committed within 12 months of placement on STMP-II. This idea is illustrated in Figure 3, which plots the cumulative count of these (proven) crimes committed by individuals in the 365 days before and after they become subject to STMP-II.

Figure 3. Cumulative count of crime before vs. after STMP-II



From Figure 3 we can see that in the 12 months prior to STMP-II, individuals in our sample collectively committed about 10,500 selected violent and property crimes.[32] In the year following placement on STMP-II, these individuals collectively committed about 8,000 crimes (a reduction of roughly 2,500 crimes). To determine whether this reduction is driven by deterrence or incapacitation, I divide the sample into cohorts based on how much time each individual spent in custody during their follow up (i.e., both sentenced and remand episodes). I then examine how much of the total reduction can be attributed to each cohort. If STMP-II is working through deterrence, we would expect the bulk of the reduction to come from individuals with little to no time in custody during their follow up. Alternatively, if STMP-II is working through incapacitation, we would expect the bulk of the reduction to come from individuals who spent most of their follow up in custody.

Table 4 defines the six (mutually exclusive) cohorts I use for this analysis.[33] The first cohort refers to individuals who spent no time in custody during their first 12 months on STMP-II. The second cohort refers to individuals who spent between 1 and 30 days in custody during their follow up. The third refers to individuals who spent between 31 and 90 days in custody. The fourth refers to individuals who spent between 91 and 180 days in custody. The fifth refers to individuals who spent between 181 and 270 days in custody. And the final cohort refers to individuals who spent between 271 and 365 days in custody during their follow up.

32 Figure A8 replicates Figure 3 for DV-STMP.
33 Figures analogous to Figures 1a and 1b for these cohorts are available for interested readers in Figure A4 of the Appendix. Such readers may also be interested in Figure A7, which plots the probability of being in custody on a given day in the 365 days before and after STMP-II.

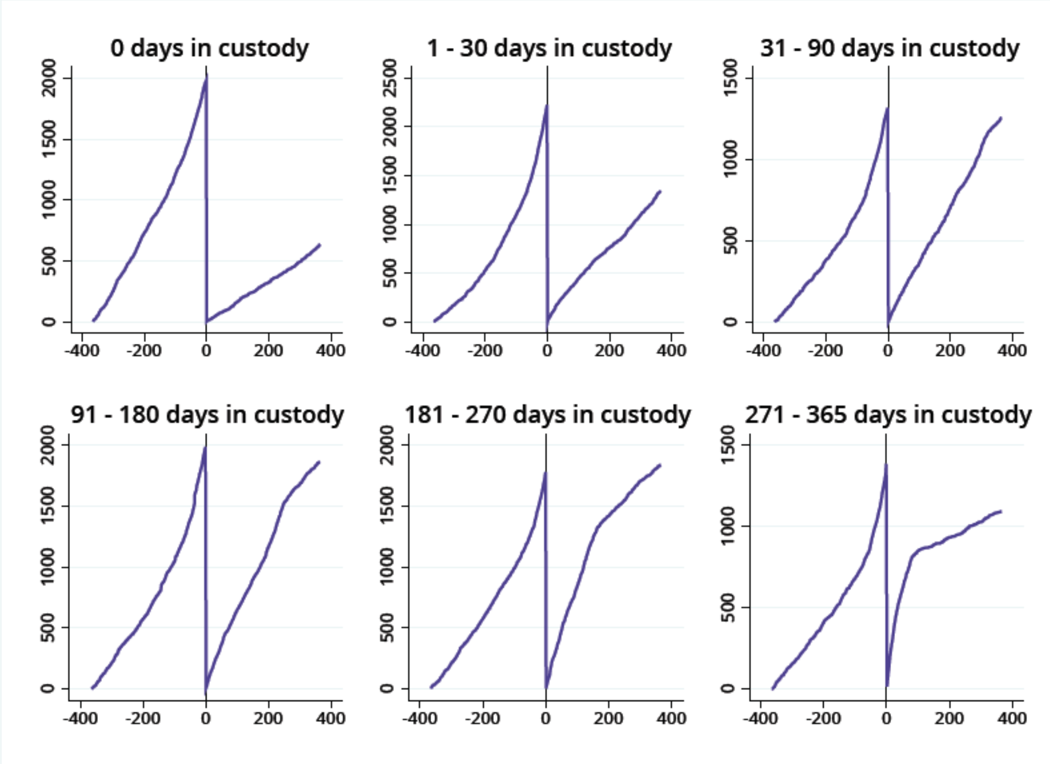
Table 4. Individuals subject to STMP-II divided into cohorts based on how much time they spent in custody during their 12 month follow up

	STMP-II		Juveniles		Aboriginal	
	Proportion	Count	Proportion	Count	Proportion	Count
	(1)	(2)	(3)	(4)	(5)	(6)
Cohort 1: No time in custody	0.376	3,806	0.293	804	0.312	1,439
Cohort 2: Between 1 and 30 days in custody	0.193	1,950	0.300	825	0.189	873
Cohort 3: Between 31 and 90 days in custody	0.104	1,049	0.147	403	0.119	550
Cohort 4: Between 91 and 180 days in custody	0.131	1,330	0.130	358	0.147	678
Cohort 5: Between 181 and 270 days in custody	0.111	1,122	0.084	232	0.131	603
Cohort 6: Between 271 and 365 days in custody	0.085	865	0.045	124	0.103	474

Figure 4 replicates Figure 3 for each cohort in Table 4.[34] From Figure 4 we can see a sharp reduction in the cumulative volume of crime committed after placement on STMP-II for individuals not placed in custody during their follow up (i.e., cohort 1). The size of the reduction appears to be in the order of about 1,400 crimes. Another 700 crimes appear to be attributable to individuals in custody for less than one month during their follow up (i.e., cohort 2). As such, taken together cohorts 1 and 2 appear to be responsible for the vast majority of the reduction observed in Figure 3. This indicates that STMP-II primarily reduces crime through deterrence. For cohorts 3 – 5, there does not appear to be any reduction in crime after placement on STMP-II. Given that individuals in these cohorts are incarcerated for anywhere between one and nine months, this suggests that such individuals offend at an extremely high rate when not in custody. This idea is further illustrated for individuals who spent at least nine months in custody during their follow up (i.e., cohort 6). Here we can see a sharp rise in the cumulative count of offences after placement on STMP-II, followed by a flattening of the curve after about 30 days. This suggests that individuals in cohort 6 offend rapidly after placement on STMP-II, before being imprisoned for the remainder of their follow up.

34 Interested readers are directed to Figures A5 and A6 in the Appendix for a graphical analysis analogous to Figures 3 and 4 for Aboriginal individuals and juveniles on STMP-II, respectively. Such readers may also be interested in Figure A9, which replicates Figure 4 for individuals subject to DV-STMP using the cohorts from Table 5.

Figure 4. Cumulative count of crime before vs. after STMP-II, by cohort



In order to formalise this analysis, I take each individual's count of selected violent and property crimes as the outcome, and then estimate an OLS regression of Equation 1.[35] These estimates, reported in Table 5, can be interpreted as the per-person reduction in selected violent and property crime after placement on STMP-II. Table 5 contains three panels. Panel A reports estimates for all participants on STMP-II. Panel B reports estimates for juveniles on STMP-II, and finally, Panel C reports estimates for Aboriginal people on STMP-II.[36]

Let's begin by focusing on Panel A. The first column reports the association between STMP-II and the total number of offences for all cohorts combined (i.e., a quantitative estimate for the reduction observed in Figure 3). We can see that in the year after placement on STMP-II, the number of crimes per-person reduces by 0.212. When aggregated across the total number of individuals subject to STMP-II, this equates to 2,132 fewer offences, which is broadly consistent with Figure 3. Now that we know STMP-II is associated with 2,132 fewer offences in total, we can re-estimate the model for each cohort to quantify how much of the reduction each cohort is responsible for.

Before proceeding it is important to make clear that the sum of the (estimated) crime reduction for each cohort (i.e., columns 2 – 7) may not equal the total crime reduction for all of the cohorts combined (i.e., column 1). This is because the coefficients reported in each column are estimated using different samples. As such, although the sum of the coefficients in columns 2 – 7 should converge on the estimate given by column 1, they may slightly over or underestimate the overall effect. In any event, the estimates from columns 2 – 7 are largely consistent with their counterparts in Figure 4 and provide us with an indication as to how STMP impacts crime.

In column 2 I restrict the estimation sample to individuals who did not spend any time in custody during their follow up (i.e., cohort 1). Here we can see that STMP-II is associated with 0.320 fewer offences per person. When aggregated across the number of people in cohort 1, this equates to about 1,205 fewer offences. This indicates that just over 50 per cent of the total reduction can be attributed purely to deterrence. Column 3 restricts the estimation sample to individuals who spent between 1 and 30 days in custody during their follow up (i.e., cohort 2). Here we can see that STMP-II is associated with 0.338 fewer

35 Interested readers are directed to Table A6 in the Appendix for analogous estimates obtained using a Negative Binomial regression.
36 Table A5 repeats this cohort analysis for individuals subject to DV-STMP.

offences per-person (657 in aggregate). Hence, together cohorts 1 and 2 account for about 87 per cent of the total reduction in offending associated with placement on STMP-II. Given that offenders in cohort 2 spend such a short duration in custody, the bulk of their reduction can likely be attributed to deterrence. Consistent with Figure 4, there is no significant relationship between STMP-II and offending for cohorts 3 – 5. Column 7 restricts the estimation sample to individuals who spent at least nine months in custody during their follow up (i.e., cohort 6). Here we can see that after placement on STMP-II, the number of offences per-person drops by 0.263 (a reduction of 227 offences). Given that individuals in this cohort spend, at minimum, three quarters of their follow up in custody, the estimate for this cohort can be attributed to incapacitation. Taken together, Panel A suggests that the crime reduction associated with STMP-II is roughly 85 per cent deterrence and 15 per cent incapacitation.

Table 5. The relationship between STMP and counts of selected violent and property crimes, by cohort

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5	Cohort 6
Panel A.	-0.212***	-0.320***	-0.338***	0.012	0.047	0.134	-0.263**
Everyone on STMP-II	(0.024)	(0.023)	(0.049)	(0.073)	(0.095)	(0.096)	(0.093)
Implied crime reduction	-2132.72	-1205.12	-656.734	-	-	-	-226.706
Observations	20,120	7,532	3,886	2,092	2,650	2,236	1,724
Adjusted R-squared	0.060	0.072	0.092	0.075	0.062	0.084	0.070
Panel B.	-0.490***	-0.769***	-0.599***	-0.040	0.239	0.646	-0.497
Juveniles	(0.074)	(0.081)	(0.117)	(0.168)	(0.282)	(0.424)	(0.795)
Implied crime reduction	-1337.7	-615.2	-491.18	-	-	-	-
Observations	5,460	1,600	1,640	804	710	460	246
Adjusted R-squared	0.033	0.093	0.062	0.019	0.021	0.022	0.012
Panel C.	-0.145***	-0.368***	-0.303***	0.091	0 296*	0 242	-0.208
Aboriginal Australians	(0.039)	(0.041)	(0.087)	(0.097)	(0.126)	(0.145)	(0.140)
Implied crime reduction	-666.42	-527.712	-263.004	-	199.5	-	-
Observations	9,192	2,868	1,736	1,096	1,348	1,200	944
Adjusted R-squared	0.061	0.071	0.076	0.086	0.076	0.107	0.072

Note. PAC = Police Area Command, robust standard errors clustered at the PAC-by-year level in parentheses, *** *p*<0.001, ** *p*<0.01, * *p*<0.05.

Panel B repeats the analysis for juveniles. From column 1 we can see that STMP-II is associated with a per-person reduction of 0.490 (1,338 fewer offences in total) for juveniles in the sample. From column 2 we can see that just under half of this reduction can be attributed to deterrence. From column 3 we can see that juveniles who spent between 1 and 30 days in custody are responsible for another one-third of the crime reduction (i.e., 0.599 fewer offences per-person, or 491 overall). There does not appear to be any significant relationship between offending and STMP-II for any of the other juvenile cohorts. Taken together, these estimates indicate that STMP-II is operating through deterrence for juveniles.

Finally, Panel C examines Aboriginal Australians on STMP-II. From column 1 we can see that STMP is associated with a per-person reduction of 0.145 (666 offences in total) across Aboriginal people in the sample. From columns 2 and 3, we can see decreases in per-person crime in the order of 0.368 and 0.303, respectively. Interestingly, in aggregate, these reductions exceed the total reduction in Aboriginal crime associated with STMP-II. One explanation is that this is the result of sampling variation. Another explanation, however, is given by the estimate in column 5. From column 5 we can see that STMP-II is associated with a per-person increase in offending of 0.296 (200 crimes in aggregate) for Aboriginal people in cohort 4.[37] This suggests that although in net terms STMP-II is associated with a decrease in offending, some Aboriginal people may be at a higher offending risk after placement on STMP-II. In any event, STMP-II is still associated with a large (net) reduction in crime committed by Aboriginal people, and the bulk of this reduction occurs through deterrence.

DISCUSSION

The Suspect Target Management Plan (STMP) is the largest and longest running offender focused policing program in NSW. At present, there are two STMP programs in operation: STMP-II, which aims to reduce general offending; and DV-STMP, which aims to reduce DV offending. This paper set out to answer three questions: first, do these programs reduce crime; second, how do these programs reduce crime; and finally, do these effects differ with regard to juveniles or Aboriginal Australians.[38] The findings from this study indicate that both programs are associated with sizable reductions in crime. With regard to the second question, while STMP-II is associated with an increased risk of imprisonment, the vast majority of any crime reduction benefit most likely occurs through deterrence, not incapacitation. DV-STMP is not associated with an increased risk of imprisonment. Therefore, DV-STMP also likely operates through deterrence. With regard to the final question, I found that the crime reduction benefit associated with STMP-II is stronger for juveniles (when compared to adults) and weaker for Aboriginal people (when compared to non-Aboriginal people). The transmission mechanism through which STMP-II affects both groups is most likely deterrence.

The present study is not, however, without its caveats. The most important of which is that the estimates do not have a causal interpretation. That is, because I am simply comparing the behaviour of individuals before and after placement on STMP, I have no way of establishing what would have happened in the absence of STMP. That said, given the increasing rate of offending prior to placement on STMP, it is possible I am underestimating the true crime reduction benefit associated with STMP. This is because there is no reason to expect that the offending of individuals placed on STMP would have declined absent the program. After all, individuals placed on STMP were already subject to the standard set of criminal justice system responses prior to placement on STMP (e.g., imprisonment, supervised orders and other sanctions) while their offending continued to rise nonetheless. With regard to STMP-II, another reason to expect that I may be underestimating the true crime reduction benefit is that I only examined a small subset of violent and property crimes (least likely to be influenced by reporting/detection bias). There is no reason to expect that STMP-II only impacts these crimes. It is, therefore, likely that STMP-II is reducing other types of crime in addition to those examined in this paper.

On the other side of this argument, however, is the fact that I examined DV offending in my analysis of DV-STMP. Recall that individuals subject to DV-STMP are explicitly notified by police following placement on the program. As such, it is possible that the reduction in (recorded) DV crime may be due to a decrease in reporting rather than offending (e.g., the perpetrator could pressure the victim not to report future instances of violence). In this case, I may be overestimating the crime reduction benefit associated with DV-STMP.

37 The increase in offending for this cohort is investigated in Table A7 of the Appendix. The increase appears to be driven by an increase in break and enter offences.

38 I did not investigate whether the effect of DV-STMP differed for juveniles or Aboriginal Australians as the sample was too small to draw any reasonable conclusions.

Another important caveat relates to how well the results from this study generalise to other settings. Rates of property crime have continued to fall in NSW since the early 2000s (Goh & Holmes, 2020). While STMP-II may have contributed to this reduction, whether STMP-II will continue to be able to generate this benefit into the future is questionable. While the results presented in this study are largely consistent with prior work on offender-focused policing programs (Braga et al., 2018) for property crime, they depart from prior work in that STMP-II does not appear to reduce violent crime for adults. The most likely explanation for this result is that the violent crimes examined in this study occur at much lower rates than in other jurisdictions (e.g., Boston or Chicago during the 1990s).

The present study makes four novel contributions to the existing body of evidence, each of which has implications for policy makers and researchers. The first is to provide some (non-causal, associative) evidence that offender-focused policing programs may work in Australia. This is an interesting possibility given that STMP differs markedly from most focused deterrence programs overseas. Focused deterrence programs typically involve working with community organisations to communicate an explicit message of deterrence. Focused deterrence programs also generally involve increasing access to social services as an adjunct to intensive policing. The fact that STMP is associated with a reduction in crime absent these features, begs the question of whether such features could further enhance the crime reduction benefit if introduced. Alternatively, it could also be the case that such features are unnecessary from a crime reduction standpoint, serving only to improve the public's perception around the equity of such programs. Exploring this question further by, for example, randomising the "carrot" as an adjunct to the "stick" in particular police jurisdictions would enable us to answer this question with minimal interruption to ongoing operations.

My second contribution is to provide evidence that the mechanism through which such programs reduce crime is not incapacitation. This suggests that such programs may be a cost-effective way to reduce crime. My third contribution is to explore how effective focused deterrence programs are in reducing DV. While the estimates reported in this paper are promising, more research is needed to answer this question definitively. As discussed earlier, the fact that DV-STMP targets are notified following placement on the program raises serious concerns around whether they are likely to retaliate against the victim, or similarly, pressure the victim to not report future violence. In this regard, careful monitoring and ideally a process evaluation of DV-STMP is certainly warranted.

My final contribution is to examine how such programs affect vulnerable groups. Although STMP is associated with a reduction in crime among Aboriginal Australians, the relationship appears to be weaker for Aboriginal people when compared to non-Aboriginal people. There is also the possibility that STMP is increasing crime for a subset of Aboriginal people, although the net effect is still an overall reduction. This suggests that the program may need to be modified for Aboriginal people to reduce the risk of any adverse outcomes. Consultation with Aboriginal elders in both the selection for, and application of the program may be one possible area for improvement. The inclusion of Aboriginal elders in the sentencing process has, for example, worked to reduce rates of recidivism and imprisonment for Aboriginal offenders (Yeong & Moore, 2020). With regard to juveniles, I found that STMP-II is associated with a significantly higher risk of imprisonment. This finding, in combination with the possibility that early engagement with, and surveillance by police, whilst subject to STMP-II may adversely influence other outcomes relevant to a young person's development (e.g., attitudes toward authority, educational achievement and mental health), begs the question of how the program can be modified to better address the needs of young people. Increased access to social services (e.g., a dedicated caseworker, mentoring, tutoring or counselling) coupled with police supervision may generate a broader benefit than police supervision alone.

A final consideration for policy makers is how police intelligence can be better utilised. An interesting finding from this study was that individuals subject to STMP were both truly distinct from other offenders in NSW and at an extremely high risk of offending. This suggests that police intelligence may be capable of identifying individuals most in need of government assistance. Such intelligence could be used, for example, to identify people for DV behavioural change programs, or alternatively, juveniles at risk of dropping out of school for education support services. If used responsibly, such information may support government in generating a safer, more equitable and prosperous society.

ACKNOWLEDGEMENTS

I would like to thank Don Weatherburn, Suzanne Poynton, Deborah Cobb-Clark, Mario Fiorini and Katrien Stevens for their feedback on earlier drafts. What a team.

REFERENCES

Angrist, J., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives, 15*(4), 69-85.

Australian Bureau of Statistics (2011). Australian and New Zealand Standard Offence Classifications. (Cat. No. 1234.0). Retrieved 16 Mar 2020 from: http://www.abs.gov.au/ausstats/abs@.nsf/mf/1234.0

Becker, G. (1968). Crime and punishment: An economic approach. *Journal of Political Economy, 76*(2), 169-217.

Blanco, C. (2018, April 19). 'Racist policing': NSW Police slammed as data reveals more than half of youth targeted by secret blacklist are Indigenous. Retrieved 9 September 2019 from *NITV*: https://www.sbs.com.au/nitv/nitv-news/article/2018/04/18/racist-policing-nsw-police-slammed-data-reveals-more-half-youth-targeted-secret

Boyle, D. J., Lanterman, J.L., Joseph E. Pascarella, J. E, & Chia-Cherng C. (2010). The Impact of Newark's Operation Ceasefire on trauma center gunshot wound admissions. *Justice Research & Policy, 12*(2), 105-123.

Braga, A. A. (2008). Pulling evers Focused Deterrence Strategies and the prevention of gun homicide. *Journal of Criminal Justice, 36*(4), 332– 343.

Braga, A. A., Apel, R., & Welsh, B. C. (2013). The spillover effects of focused deterrence on gang violence. *Evaluation Review, 37*(3-4), 314–342.

Braga., A. A., Hureau, D. M., & Papachristos, A. V. (2014). Deterring Gang-Involved Gun Violence: Measuring the Impact of Boston's Operation Ceasefire on Street Gang Behaviour. *Journal of Quantitative Criminology, 30*(1), 113–139.

Braga, A. A., Kennedy, D. M., Waring, E. J., & Piehl, A. M. (2001). Problem-oriented policing, deterrence, and youth violence: An evaluation of Boston's Operation Ceasefire. *Journal of Research in Crime & Delinquency, 38*(3), 195–225.

Braga, A. A., Pierce, G. L., McDevitt, J., Bond, B. J., and Cronin, S. (2008). The strategic prevention of gun violence among gang-involved offenders. *Justice Quarterly, 25*(1), 132–162.

Braga, A. A., Weisburd, D., & Turchan, B. (2018). Focused deterrence strategies and crime control: An updated systematic review and meta-analysis of the empirical evidence. *Criminology & Public Policy, 17*(1), 205-250.

Chalfin, A., & McCrary, J. (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature, 55*(1), 5-48.

Chalfin, A., & McCrary, J. (2018). Are U.S. cities underpoliced? Theory and evidence. *The Review of Economics and Statistics, 100*(1), 167-186.

Circo, G., Krupa, J. M., McGarrell, E., & DeBiasi, A. (2020). The individual-level deterrent effect of "call-in" meetings on time to re-arrest. *Crime & Delinquency, 66*(11), 1630-1651.

Cobb-Clark, D., Kettlewell, N., Schurer, S., & Silburn, S. (2018). The Effect of Quarantining Welfare on School Attendance in Indigenous Communities. *IZA Discussion paper series*, no. 11514.

Corman, H. & Mocan, N. H.N. (2000). A Time-Series Analysis of Crime, Deterrence, and Drug Abuse in New York City. *American Economic Review, 90*(3), 584-604.

Corsaro, N., Brunson, R., & McGarrell, E. (2010). Evaluating a policing strategy intended to disrupt an illicit street-level drug market. *Evaluation Review, 34(*6), 513-548.

Corsaro, N., Brunson, R., & McGarrell, E. (2013). Problem-Oriented Policing and Open-Air Drug Markets: Examining the Rockford Pulling Levers Strategy. *Crime & Delinquency, 59*(7) 1085-1107.

Corsaro, N., Hunt, E. D., Hipple, N. K., & McGarrell, E. F. (2012). The impact of drug market pulling levers policing on neighborhood violence. *Criminology & Public Policy, 11*(2), 167–199.

Engel, R. S., Corsaro, N., & Tillyer, M. S. (2013). Reducing gang violence using focused deterrence: Evaluating the Cincinnati Initiative to Reduce Violence (CIRV). *Justice Quarterly, 30*(3), 403-439

Evans, W. N. & Owens, E. G. (2007). COPS and crime. *Journal of Public Economics, 91*(1-2), 181-201.

DeAngelo, G., & Hansen, B. (2014). Life and death in the fast lane: Police enforcement and traffic fatalities. *American Economic Journal: Economic Policy, 6*(2), 231-257

Di Tella, R. & Schargodsky, E. (2004). Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *American Economic Review, 94*(1), 115-133.

Goh, D. & Holmes, J. (2020). *An update of long-term trends in property and violent crime in New South Wales: 1990-2019* (Bureau Brief No. 144). Sydney: NSW Bureau of Crime Statistics and Research.

Grunwald, B., & Papachristos, A. (2017). Project Safe Neighborhoods in Chicago: Looking back a decade later. *Journal of Criminal Law and Criminology, 107*(1), 131-160.

Hoynes, H.W., & Schanzenbach, D. W. (2009). Consumption responses to in-kind transfers: Evidence from the introduction of the Food Stamp Program. *American Economic Journal: Applied Economics, 1*(4), 109–139

Judicial Commission of NSW (2020). About law part codes. Retrieved 18 Jun 2020 from Judicial Commission of NSW website: https://lawcodes.judcom.nsw.gov.au/

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis, 27*(4), 1–20.

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

Levitt, S. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *American Economic Review, 87*(3), 270-290.

Machin, S., & Marie, O. (2011). Crime and police resources: The Street Crime Initiative. *Journal of the European Economic Association, 9*(4), 678-701.

Marvell, T., & Moody, C. E. (1996). Specification problems, police levels, and crime rates. *Criminology, 34*(4), 609-646.

McGarrell, E. F., Chermak, S., Wilson, J., & Corsaro, A. (2006). Reducing homicide through a 'Lever-Pulling' Strategy. *Justice Quarterly, 23*, 214–231.

McGowan, M. (2017, November 11). More than 50% of those on secretive NSW police blacklist are Aboriginal. *The Guardian*. Retrieved 9 September 2019 from: https://www.theguardian.com/australia-news/2017/nov/11/more-than-50-of-those-on-secretive-nsw-police-blacklist-are-aboriginal

O'Mallon, F. (2017, November 11). NSW Police visit homes of people on secret watchlist without cause. *The Sydney Morning Herald*. Retrieved 9 September 2019 from: https://www.smh.com.au/national/nsw/nsw-police-visit-homes-of-people-on-secret-watchlist-without-cause-20171107-gzgcwg.html

Owens, E. G. (2013). COPS and Cuffs. In Cook. P., Machin, S., Marie, O, & Mastrobuoni, G (1st Ed.), Lessons from the Economics of Crime: *What Reduces Offending?* (pp.17-4 44). Cambridge, Massachusetts: The MIT Press.

Papachristos, A., Meares, T., & Fagan, J. (2007). Attention felons: Evaluating Project Safe Neighborhoods in Chicago. *Journal of Empirical Legal Studies, 4*(2), 223– 272.

Rosenfeld, R., Fornango, R., & Baumer, E. (2005). Did Ceasefire, Compstat, and Exile reduce homicide? *Criminology & Public Policy, 4*(3), 419–450.

Saunders, J., Lundberg, R., Braga, A. A., Ridgeway, G., & Miles, J. (2015). A synthetic control approach to evaluating place-based crime interventions. *Journal of Quantitative Criminology, 31*(3), 413–434.

Sentas, V., & Pandolfini, C. (2017). *Policing Young People in NSW: A Study of the Suspect Targeting Management Plan. A Report of the Youth Justice Coalition NSW.* Sydney: Youth Justice Coalition NSW.

Shi, L. (2009). The limit of oversight in policing: Evidence from the 2001 Cincinnati riot. *Journal of Public Economics, 93*(1-2), 99-113.

Shoebridge, D. (2018). *Media Release: NSW Police secret blacklist is nothing more than racist policing.* Retrieved 9 September 2019 from: https://davidshoebridge.org.au/2018/03/15/media-release-nsw-police-secret-blacklist-is-nothing-more-than-racist-policing/

Weisburd, D., Majmundar, M., Aden, H., Braga, A., Bueermann, J., Cook, P., Goff, P., Harmon, R., Haviland, A., Lum, C., Manski, C., Mastrofski, S., Meares, T., Nagin, D., Owens, E., Raphael. S., Ratcliffe, J., & Tyler, T. (2019). Proactive policing: A summary of the report of the National Academies of Sciences, Engineering, and Medicine. Asian Journal of Criminology, 14, 145-177.

Williams, D.J., Currie, D., Linden, W., & Donnelly, P.D. (2014). Addressing gang-related violence in Glasgow: A preliminary pragmatic quasi-experimental evaluation of the Community Initiative to Reduce Violence (CIRV). Aggression and Violent Behavior, 19(6), 686–691.

Yeong, S. (2019). *The effect of police on crime and arrests: Are police deterring or incapacitating criminals?* (Crime and Justice Bulletin No. 223). Sydney: NSW Bureau of Crime Statistics and Research.

Yeong, S., & Moore, E. (2020). *Circle Sentencing, incarceration and recidivism* (Crime and Justice Bulletin No. 226). Sydney: NSW Bureau of Crime Statistics and Research.

APPENDIX

An interesting feature of STMP is the fact that an individual exerts no influence over whether they are placed on STMP (other than through their offending behaviour). This lends credibility to the use of a matching strategy, which often fail in settings where individuals have a clear incentive to opt-into the program (e.g., in job training programs). The question, therefore, is whether there exists a subset of individuals in ROD, who resemble individuals on STMP, that were not subject to STMP for reasons unrelated to their risk of offending. Such individuals could, for example, include offenders prone to drug crime that reside within the jurisdiction of a PAC focusing on violent crime. The identifying assumption is that if this individual instead resided within the jurisdiction of a PAC focusing on drug crime, he would be subject to STMP.

In order to investigate this idea, I reorganise the data into a quarterly individual level panel (i.e., one row per individual per quarter-year), and then for each observation, calculate the values of the control variables (from Table 1) as of the first day of each quarter.[39] I then match individuals who were observably similar on the first day of the quarter that the treated unit began STMP. For example, if an individual begins STMP on 15 February 2010, I match this individual to one who was observably similar on 1 January 2010 (i.e., the first day of 2010-Q1) using the controls in Table 1.

The matching algorithm used to generate these estimates involves the following six steps:

1. Limit the sample to a given quarter-year (e.g., 2010-Q1).
2. Retain observations that either: began STMP within the given quarter-year or were never subject to STMP.[40]
3. Use Coarsened Exact Matching (CEM) to further limit the sample to the subset of treatment-control observations within the area of common support.
4. Use Propensity Score Matching (PSM) to obtain the best possible (1:1) match between each treatment and control unit.
5. Recalculate all variables to the day that the treated unit began STMP.[41]
6. Repeat steps 1 – 5 for each quarter-year between 2005-Q2 and 2018-Q3.

This matching algorithm has several advantages over using CEM or PSM alone. As outlined by King and Nielsen (2019) the use of CEM before PSM safeguards against extrapolations made in PSM that can lead to model dependant inferences. King and Nielsen (2019) also argue that PSM, used in conjunction with CEM, is potentially better than CEM alone in situations with a large degree of imbalance between groups. Said differently, PSM (after CEM) works well in circumstances where causal inferences are least likely because the treatment and control groups are so different. Another advantage to pre-processing the data using CEM is that CEM is computationally efficient, which is of practical importance given that I have, for each quarter-year, over 1.3 million potential control units.

39 For example, each individual's age on 1 January 2010 for 2010-Q1, 1 April 2010 for 2010-Q2 and so on.
40 Potential control units that were in custody in the first day of the quarter are excluded from donor pool.
41 That is, because of variables are indexed to the first day of a given quarter-year, I need to re-index these variables to the day that the treated unit began STMP. For example, within a given pair, if the treated unit began STMP on 1 February 2010, then I recalculate both the treated and control unit's control variables (e.g., number of prior offences) as of 1 February 2010 (instead of the first day of 2010-Q1). I also index my outcome variables (e.g., reoffend within 12 months) to the first day that the treated unit began STMP within each pair.

Table A1 reports the results from a balance test between groups for STMP-II and DV-STMP, in Panels A and B, respectively.

Table A1. Comparison of matched treatment and control groups, STMP-II and DV-STMP

	Matched control group			Treatment group			Difference	
	N	Mean	Std. Dev.	N	Mean	Std. Dev.	Estimate	Std. Err
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. STMP-II								
Age	9,349	27.828	9.652	9,461	26.312	10.150	-1.515***	(0.144)
Age at first CJS contact	9,349	18.125	6.467	9,462	17.880	6.821	-0.244*	(0.097)
Male	9,349	0.914	0 280	9,463	0.918	0.274	0.004	(0.004)
Aboriginal	9,349	0.508	0.500	9,464	0.457	0.498	-0.052***	(0.007)
Prior court appearances	9,349	10.924	8.608	9,465	10.077	7.464	-0.847***	(0.118)
Prior violent offences	9,349	5.023	5.692	9,466	4.659	5.203	-0.364***	(0.080)
Prior weapon offences	9,349	0.506	1 370	9,467	0.557	1.359	0.052**	(0.012)
Prior drug offences	9,349	1.764	2.952	9,468	1.722	2.805	-0.043	(0.042)
Prior community orders	9,349	4.314	3 283	9,469	4.805	3.069	0.491***	(0.046)
Prior YJCs and cautions	9,349	0.714	1.191	9,470	0.707	1.180	-0.007	(0.017)
Prior prison sentences	9,349	2.062	3.455	9,471	2.068	3.280	0.005	(0.049)
Panel B. DV-STMP								
Age	960	35.660	9.431	969	35.209	10.325	-0.451	(0.450)
Age at first CJS contact	960	21 377	7.861	970	21.627	8.975	0.250	(0.384)
Male	960	0.917	0 277	971	0.920	0.272	0.003	(0.013)
Aboriginal	960	0.397	0.490	972	0.379	0.485	-0.018	(0.022)
Prior court appearances	960	11.986	9 284	973	11.312	8.412	-0.675	(0.403)
Prior violent offences	960	9.123	8.116	974	9.120	7.632	-0.003	(0.359)
Prior weapon offences	960	0.602	2.756	975	0.516	1.312	-0.086	(0.098)
Prior drug offences	960	1.819	3.116	976	1.576	2.381	-0.243	(0.126)
Prior DV offences	960	7.383	10.589	977	8.279	6.993	0.895*	(0.409)
Prior community orders	960	4.300	3 262	978	4.495	2.985	0.195	(0.142)
Prior YJCs and cautions	960	0.455	1.154	979	0.434	1.017	-0.021	(0.050)
Prior prison sentences	960	2.045	3.610	980	1.937	3.264	-0.108	(0.157)

Note. N = Observations, YJC = youth justice conference, CJS = criminal justice system, robust standard errors in parentheses, *** $p<0.001$, ** $p<0.01$, * $p<0.05$.

From Table A1 we can see that, despite the extremely large donor pool of potential controls (over 1.3 million for each quarter-year), the matched groups are not statistically or practically equivalent to their respective treatment groups. Individuals subject to STMP are, on average, younger, less likely to be Aboriginal (for STMP-II), have fewer court appearances, violent offences, offences involving the use of a weapon and prison sentences. This would suggest that the control group is of a higher risk than individuals subject to STMP.

However, in Table A2 I report estimates from several regressions comparing the recidivism rates between these groups. Interestingly, these regressions indicate that the reverse is actually true: individuals subject to STMP offend at much higher rates than their matched counterparts.[42] One explanation for this finding is that there is some form of unobserved heterogeneity that matching cannot address. For example, known criminal associations, addiction issues and police intelligence are important unobserved factors likely to influence program participation.

42 Interested readers are directed to Figure A1 for figures that plot the daily probability of crime and a custodial episode for the matched control group.

Table A2. Matched regressions

	(1) Naive	(2) Controls	(3) PAC fixed effects
Panel A.	0.258***	0.225***	0.241***
STMP-II on selected violent or property crime	(0.011)	(0.009)	(0.013)
Observations	18,486	18,486	18,486
Adjusted R-squared	0.0900	0.154	0.154
Panel B.	0.303***	0.288***	0.324***
STMP-II on selected violent or property crime	(0.021)	(0.019)	(0.026)
Observations	1,906	1,906	1,906
Adjusted R-squared	0.114	0.158	0.157

Note. PAC = Police Area Command, FE = fixed effects, robust standard errors in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Table A3. DV-STMP and specific types of DV offences

	(1) Assault	(2) Assault excl stalking	(3) Sexual offences	(4) Property damage	(5) Justice proce-dures offences	(6) Serious assault, resulting in injury	(7) Serious assault, not resulting in injury	(8) Common assault	(9) Stalking
DV-STMP	-0.293***	-0.228***	-0.001	-0.115***	-0.194***	-0.076***	-0.002	-0.176***	-0.173***
	(0.023)	(0.021)	(0.002)	(0.013)	(0.020)	(0.014)	(0.001)	(0.018)	(0.019)
Observations	2,050	2,050	2,050	2,050	2,050	2,050	2,050	2,050	2,050
Adjusted R-squared	0.117	0.092	0.007	0.066	0.073	0.025	0.015	0.075	0.067

Note. PAC = Police Area Command, FE = fixed effects, robust standard errors clustered at the PAC-by-year level in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Table A4. Descriptive statistics for outcome variables

	Before STMP		First day on STMP		Difference	
	Mean	Std Dev	Mean	Std Dev	Estimate	Std. Err
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Everone on STMP-II						
At least one selected violent or property crime within 12 months excluding a three month interval on either side of the STMP start date	0.272	0.445	0.232	0.422	-0.041***	(0.006)
At least one selected violent or property crime within 24 months excluding a three month interval on either side of the STMP start date	0.417	0.493	0.357	0.479	-0.060***	(0.007)
At least one selected violent or property crime within 24 months excluding a six month interval on either side of the STMP start date	0.356	0.479	0.307	0.461	-0.049***	(0.006)
At least one selected violent or property crime within 24 months	0.522	0.500	0.433	0.495	-0.090***	(0.007)
At least one sentence of imprisonment within 12 months	0.189	0 391	0.303	0.460	0.115***	(0.006)
At least one selected violent crime within 12 months	0.050	0 218	0.046	0.209	-0.004	(0.003)
Count of selected violent crime within 12 months	0.071	0 367	0.064	0.364	-0.007	(0.005)
At least one selected property crime within 12 months	0.393	0.489	0.312	0.463	-0.082***	(0.007)
Count of selected property crime within 12 months	1.012	2.073	0.731	1.702	-0.281***	(0.027)
At least one selected violent or property crime within 12 months	0.418	0.493	0.335	0.472	-0.083***	(0.007)
Count of selected violent or property crime within 12 months	1.083	2.115	0.795	1.765	-0.288***	(0.027)
Panel B. Juveniles on STMP-II						
At least one selected violent or property crime within 12 months excluding a three month interval on either side of the STMP start date	0.454	0.498	0.382	0.486	-0.072***	(0.013)
At least one selected violent or property crime within 24 months excluding a three month interval on either side of the STMP start date	0.591	0.492	0.551	0.497	-0.040**	(0.013)
At least one selected violent or property crime within 24 months excluding a six month interval on either side of the STMP start date	0.490	0.500	0.491	0.500	0.001	(0.014)
At least one selected violent or property crime within 24 months	0.744	0.437	0.629	0.483	-0.115***	(0.012)
At least one sentence of imprisonment within 12 months	0.131	0 338	0.267	0.442	0.135***	(0.011)
At least one selected violent crime within 12 months	0.133	0 339	0.097	0.295	-0.036***	(0.009)
Count of selected violent crime within 12 months	0.191	0.571	0.131	0.484	-0.060***	(0.014)
At least one selected property crime within 12 months	0.594	0.491	0.459	0.498	-0.135***	(0.013)
Count of selected property crime within 12 months	1.800	2.659	1 231	2.351	-0.570***	(0.068)
At least one selected violent or property crime within 12 months	0.654	0.476	0.498	0.500	-0.156***	(0.013)
Count of selected violent or property crime within 12 months	1.991	2.701	1 362	2.438	-0.629***	(0.069)

Table A4. Descriptive statistics for outcome variables (continued)

	Before STMP		First day on STMP		Difference	
	Mean	Std Dev	Mean	Std Dev	Estimate	Std. Err
	(1)	(2)	(3)	(4)	(5)	(6)
Panel C. Aboriginal Australians						
At least one selected violent or property crime within 12 months excluding a three month interval on either side of the STMP start date	0.319	0.466	0.286	0.452	-0.033***	(0.010)
At least one selected violent or property crime within 24 months excluding a three month interval on either side of the STMP start date	0.481	0.500	0.437	0.496	-0.044***	(0.010)
At least one selected violent or property crime within 24 months excluding a six month interval on either side of the STMP start date	0.417	0.493	0.377	0.485	-0.040***	(0.010)
At least one selected violent or property crime within 24 months	0.585	0.493	0.520	0.500	-0.064***	(0.010)
At least one sentence of imprisonment within 12 months	0.224	0.417	0.349	0.477	0.125***	(0.009)
At least one selected violent crime within 12 months	0.052	0 223	0.058	0.234	0.006	(0.005)
Count of selected violent crime within 12 months	0.072	0 346	0.086	0.424	0.014	(0.008)
At least one selected property crime within 12 months	0.445	0.497	0.379	0.485	-0.066***	(0.010)
Count of selected property crime within 12 months	1.173	2.035	0.937	1.982	-0.236***	(0.042)
At least one selected violent or property crime within 12 months	0.468	0.499	0.405	0.491	-0.063***	(0.010)
Count of selected violent or property crime within 12 months	1.245	2.079	1.023	2.056	-0.222***	(0.043)
Panel D. Everyone on DV-STMP						
At least one DV crime within 12 months excluding a three month interval on either side of the STMP start date	0.480	0.500	0.279	0.449	-0.200***	(0.021)
At least one DV crime within 24 months excluding a three month interval on either side of the STMP start date	0.603	0.489	0.363	0.481	-0.240***	(0.021)
At least one DV crime within 24 months excluding a six month interval on either side of the STMP start date	0.447	0.497	0.296	0.457	-0.152***	(0.021)
At least one selected violent or property crime within 24 months	0.797	0.403	0.486	0.500	-0.310***	(0.020)
At least one sentence of imprisonment within 12 months	0.220	0.414	0.265	0.441	0.045*	(0.019)
At least one selected violent crime within 12 months	0.007	0.082	0.005	0.070	-0.002	(0.003)
Count of selected violent crime within 12 months	0.007	0.082	0.005	0.070	-0.002	(0.003)
At least one selected property crime within 12 months	0.087	0 281	0.075	0.263	-0.012	(0.012)
Count of selected property crime within 12 months	0.121	0.468	0.122	0.556	0.001	(0.023)
At least one selected violent or property crime within 12 months	0.091	0 288	0.079	0.270	-0.013	(0.012)
Count of selected violent or property crime within 12 months	0.127	0.484	0.126	0.563	-0.001	(0.023)
At least one DV crime within 12 months	0.721	0.449	0.418	0.494	-0.303***	(0.021)
Count of DV crime within 12 months	2.482	3 361	1.076	1.799	-1.407***	(0.119)

Note. Robust standard errors in parentheses, *** p<0.001, ** p<0.01, * p<0.05.

Table A5. Change in counts of DV crime after DV-STMP, by cohorts described in Table 4

	(1) Full sample	(2) Cohort 1	(3) Cohort 2	(4) Cohort 3	(5) Cohort 4	(6) Cohort 5	(7) Cohort 6
DV-STMP on count of DV crime	-1.300***	-1.645***	-0.623**	-0.467	0.100	-0.476	-1.936*
	(0.102)	(0.112)	(0.207)	(0.375)	(0.500)	(0.335)	(0.885)
Implied crime reduction	-1332.5	-687.61	-118.99				-75.5
Observations	2,050	836	382	240	328	186	78
Adjusted R-squared	0.082	0.224	0.098	0.057	0.010	0.055	0.001

Note. PAC = Police Area Command, FE = fixed effects, robust standard errors clustered at the PAC-by-year level in parentheses, *** *p*<0.001, ** *p*<0.01, * *p*<0.05.

Table A6. Maximum likelihood robustness checks

	(1) Prob (Crime)	(2) Prob (Imprisonment)	(3) Count (Crime)
Panel A. STMP-II on everyone	-0.060***	0.096***	-0 225***
	(0.006)	(0.006)	(0.025)
Observations	19,784	19,538	20,120
Specification	Probit	Probit	Negative binomial
Pseudo R-sqauared	0.139	0.130	0.079
Area under the receiver operating characteristic curve	0.743	0.747	
Panel B. STMP-II on juveniles	-0.152***	0.105***	-0.658***
	(0.014)	(0.013)	(0.068)
Observations	5,086	4,152	5,460
Specification	Probit	Probit	Negative binomial
Pseudo R-sqauared	0.142	0.230	0.082
Area under the receiver operating characteristic curve	0.743	0.819	
Panel C. STMP-II on Aboriginal Australians	-0.044***	0.109***	-0.165***
	(0.010)	(0.009)	(0.039)
Observations	8,868	8,618	9,192
Specification	Probit	Probit	Negative binomial
Pseudo R-sqauared	0.141	0.147	0.085
Area under the receiver operating characteristic curve	0.745	0.756	
Panel D. DV-STMP on everyone	-0.306***	0.007	-1.452***
	(0.021)	(0.020)	(0.118)
Observations	1,986	1,828	2,050
Specification	Probit	Probit	Negative binomial
Pseudo R-sqauared	0.169	0.137	0.064
Area under the receiver operating characteristic curve	0.768	0.751	

Note. This table reports average marginal effects from various Maximum Likelihood regressions, standard errors obtained using the Delta method in parentheses, *** *p*<0.001, ** *p*<0.01, * *p*<0.05.

Table A7. Counts of offences per-person for Aboriginal people in cohort 4

	(1) GBH	(2) Homicide	(3) Robbery	(4) Break and enter	(5) Theft
STMP-II	0.005	-0.001	0.028	0.128	0.136
	(0.006)	(0.001)	(0.021)	(0.068)	(0.086)
Observations	1,348	1,348	1,348	1,348	1,348
Adjusted R-squared	0.004	0.001	0.021	0.059	0.040

Note. GBH = assault occasioning grievous bodily harm, robust standard errors clustered at the PAC-by-year level in parentheses,
*** *p*<0.001, ** *p*<0.01, * *p*<0.05.

Figure A1. Daily rates of offending and custody for the matched control group described in Table A1

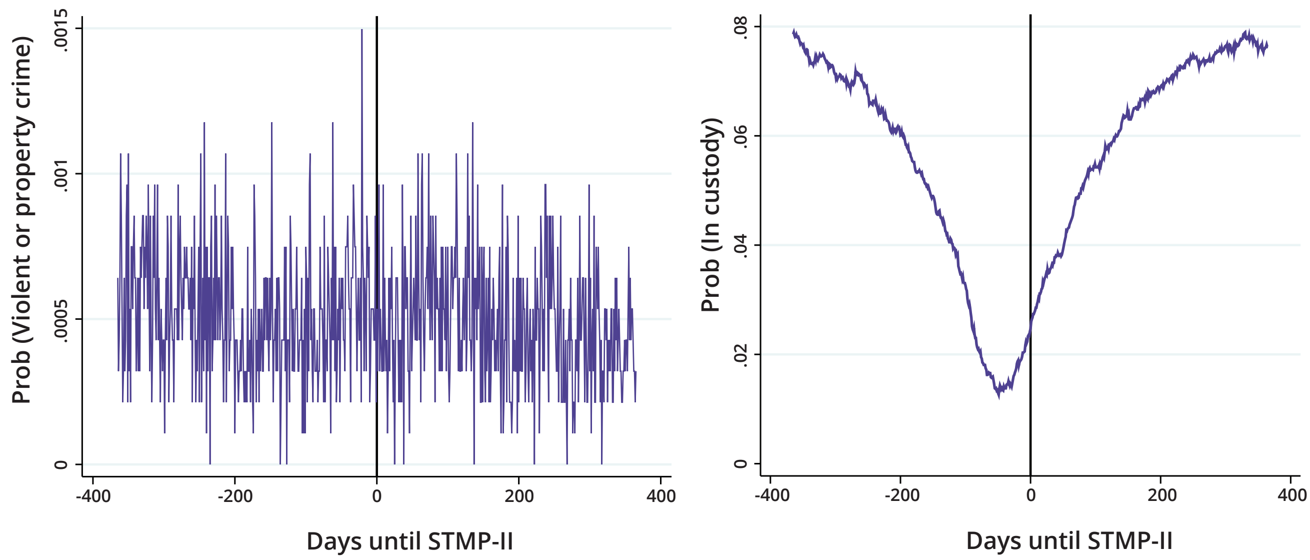


Figure A2. 730 days of offending and custody

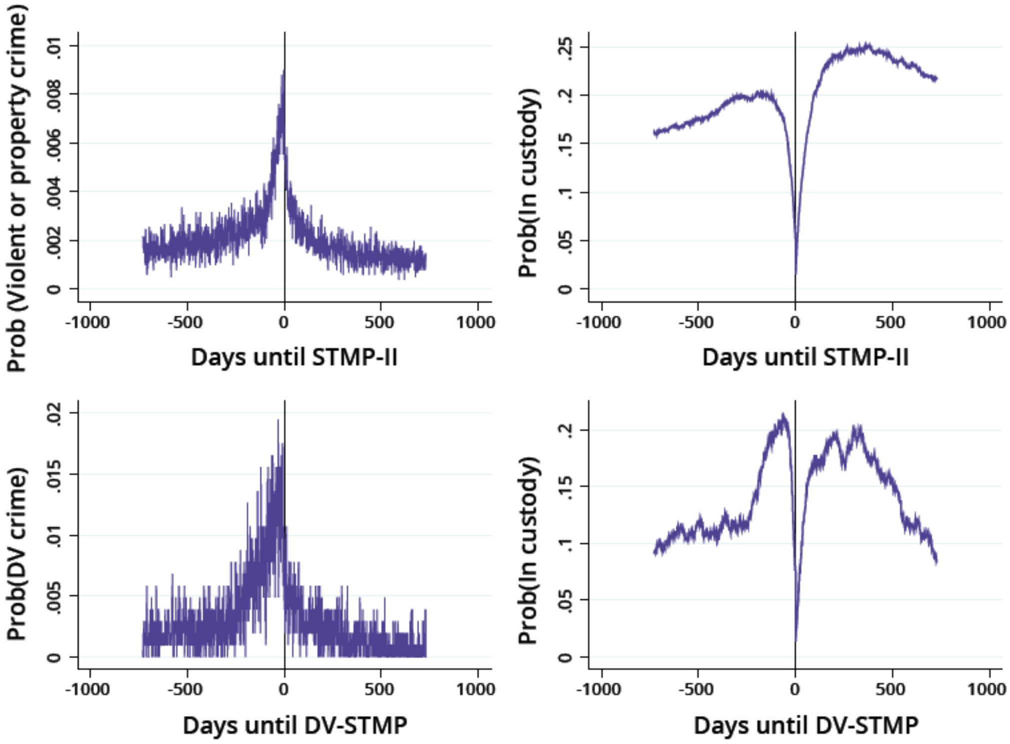


Figure A3. Daily rates of offending and custody for juveniles and Aboriginal Australians on STMP-II

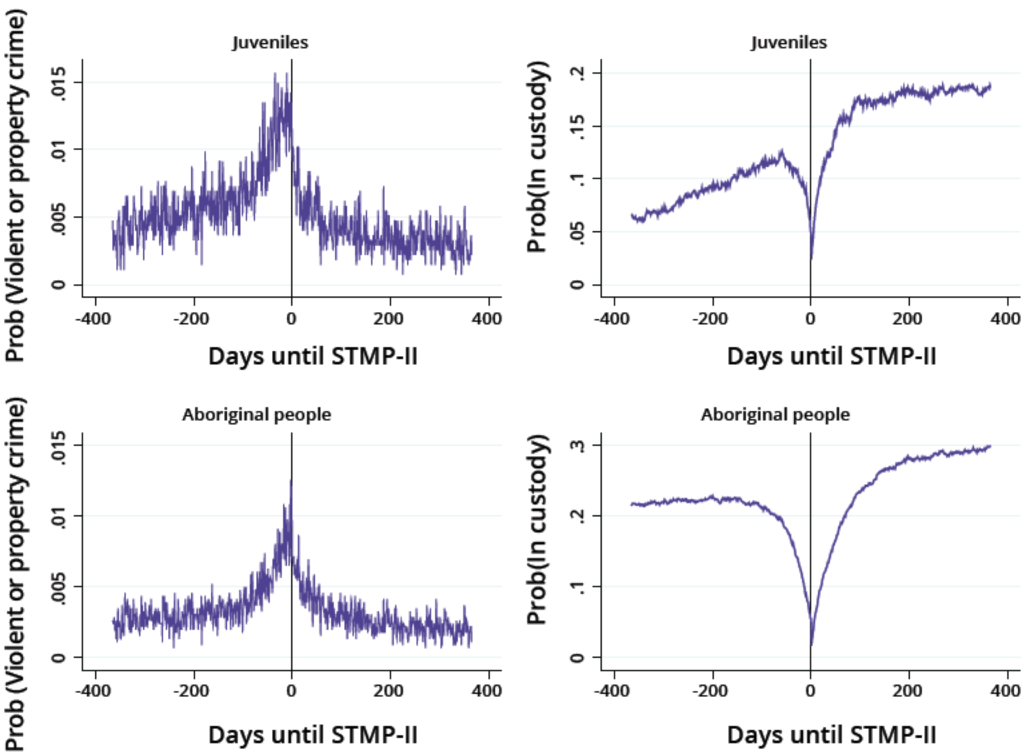


Figure A4. Daily probability of offending before and after STMP-II, by cohorts described in Table 4

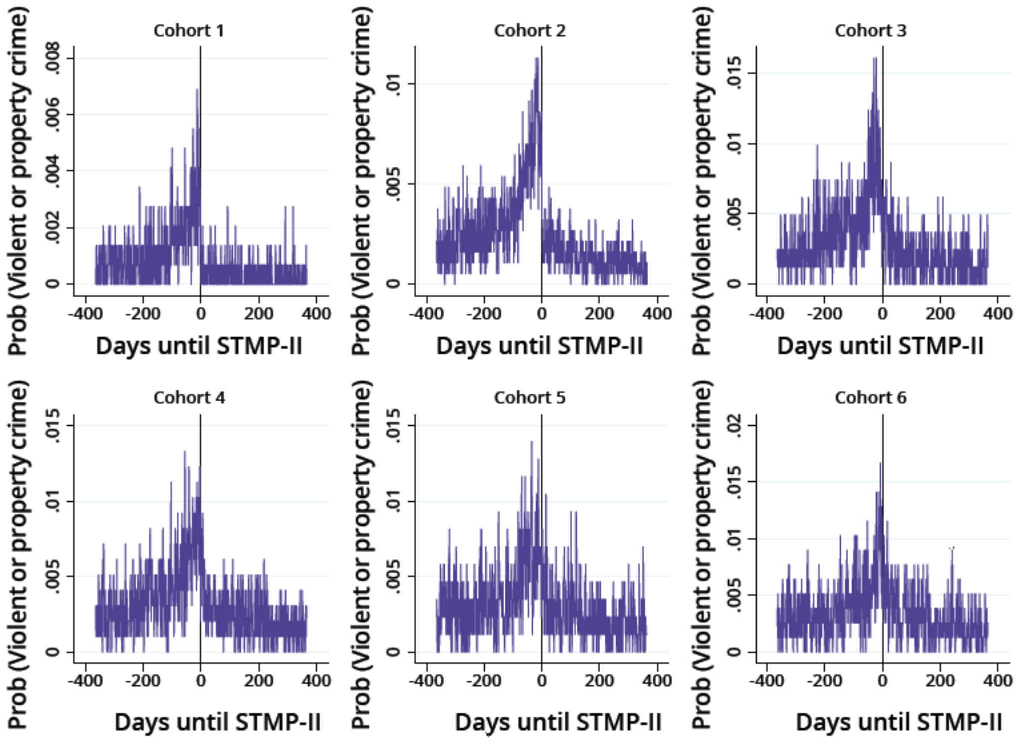


Figure A5. Counts of Aboriginal offending before and after STMP-II, by cohorts described in Table 4

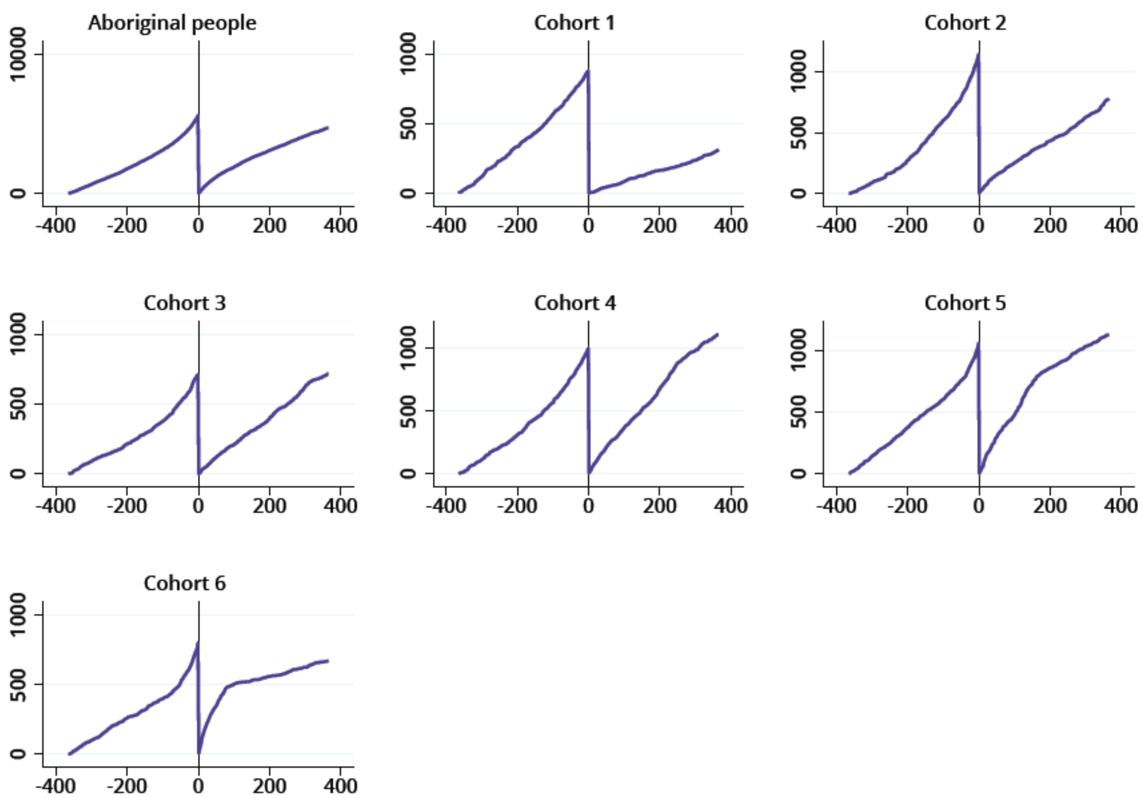


Figure A6. Counts of juvenile offending before and after STMP-II, by cohorts described in Table 4

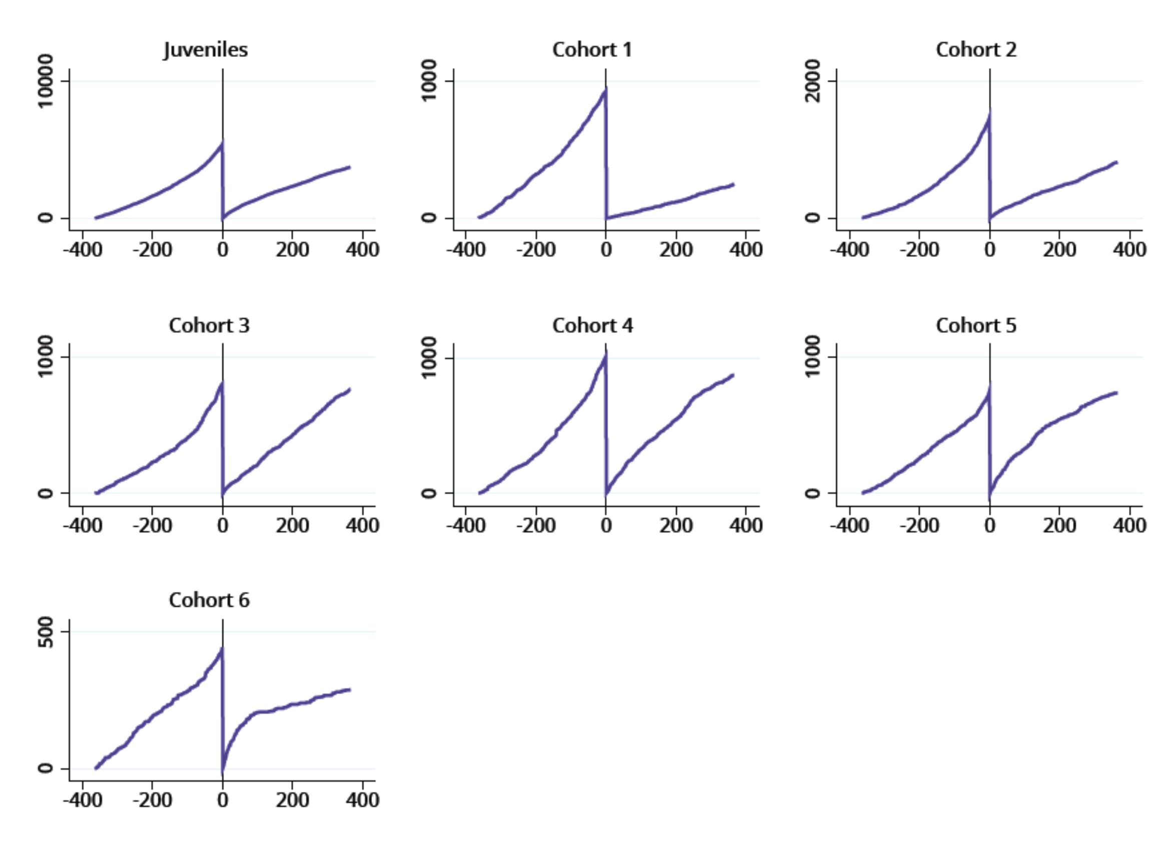


Figure A7. Daily rates of custody before and after STMP-II by cohort

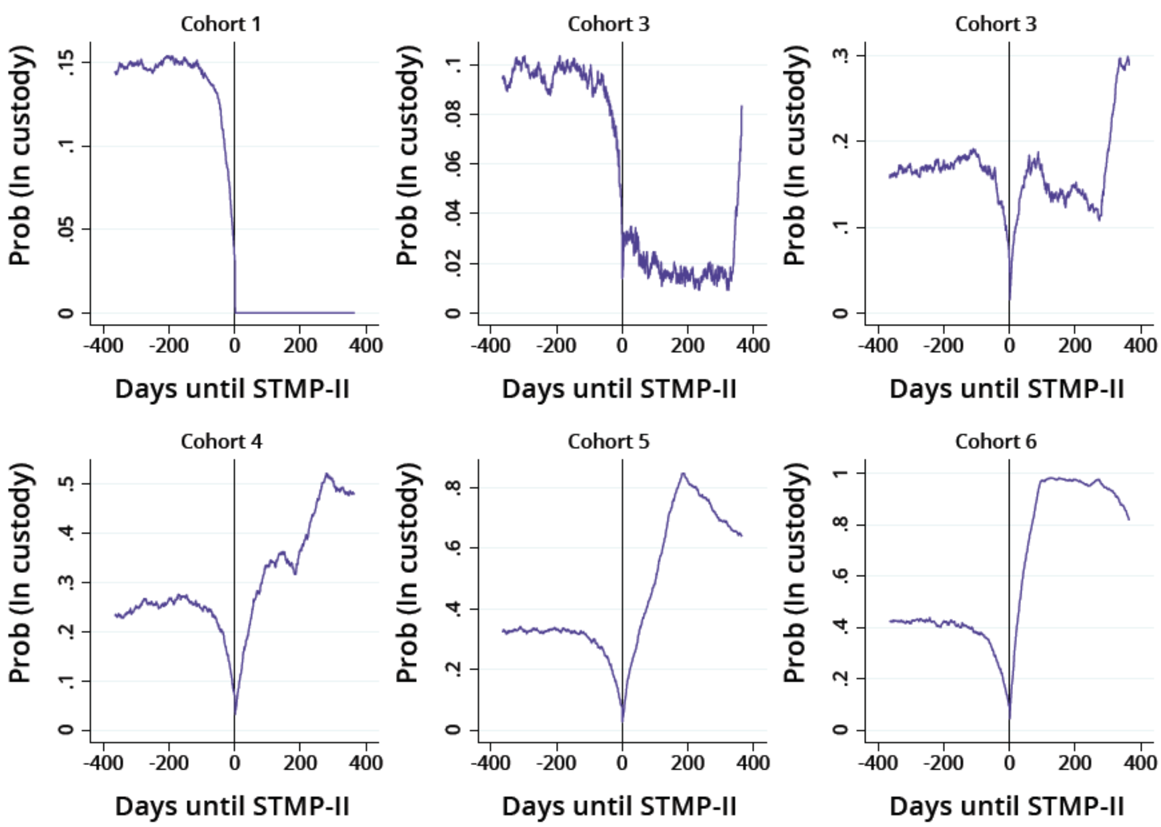


Figure A8. Counts of offences before and after DV-STMP

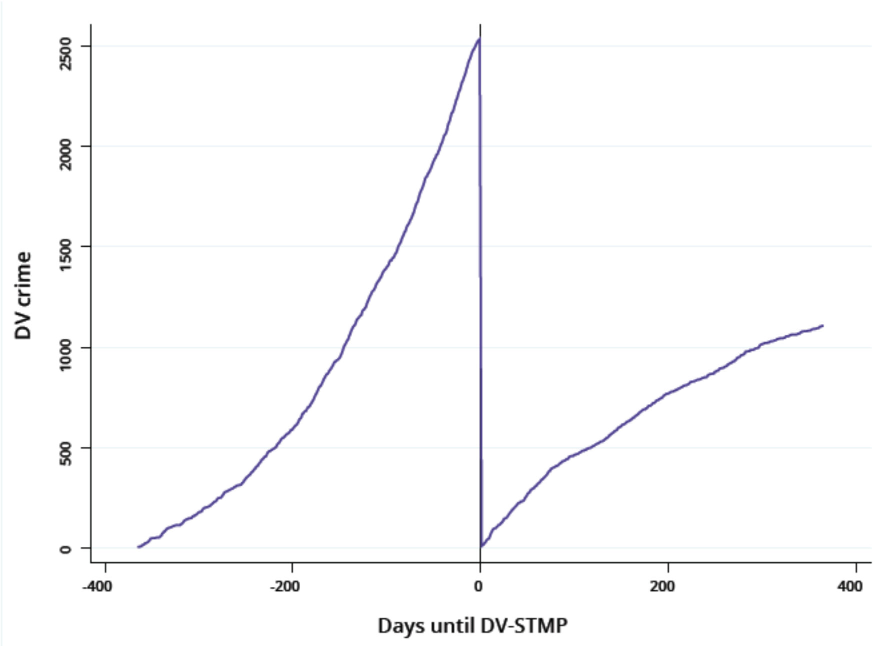
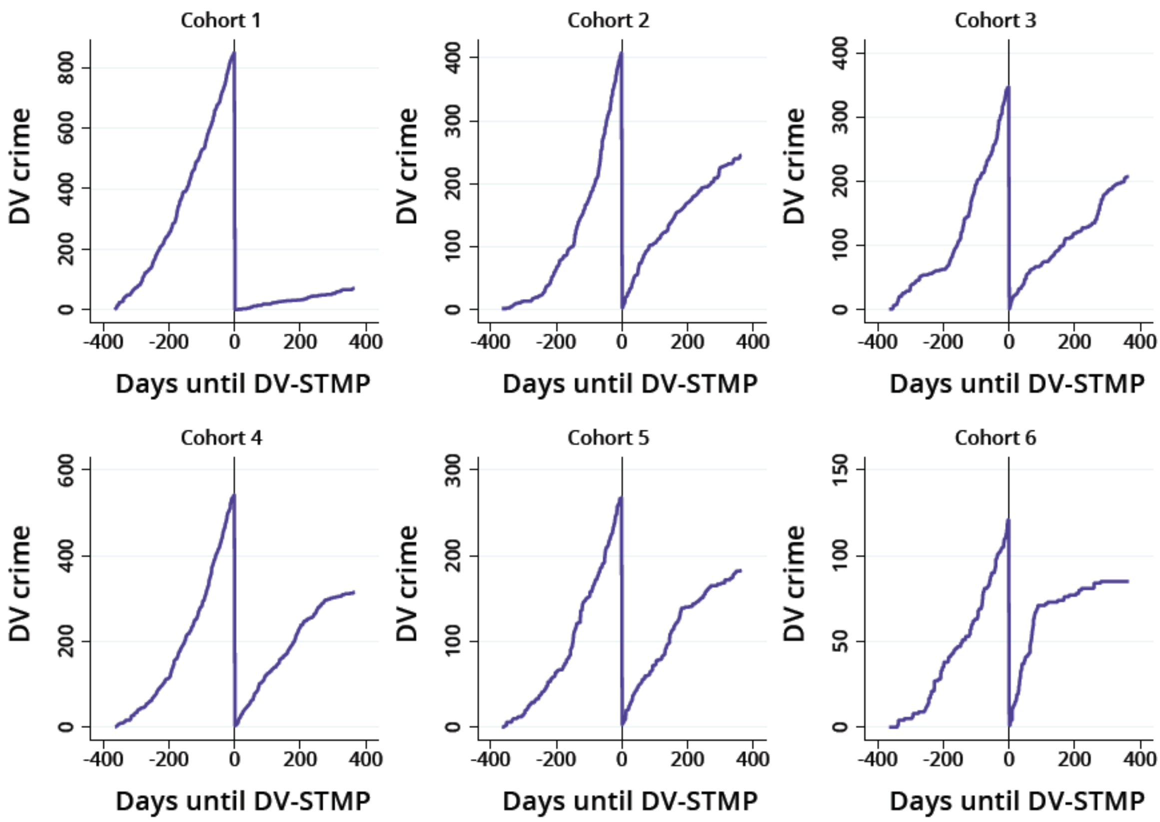


Figure A9. Counts of offences before and after DV-STMP, by cohorts described in Table 4



NSW BUREAU OF CRIME STATISTICS AND RESEARCH - LEVEL 1, HENRY DEANE BUILDING, 20 LEE STREET, SYDNEY 2000
bcsr@justice.nsw.gov.au · www.bocsar.nsw.gov.au · Ph: (02) 8346 1100 · Fax: (02) 8346 1298
ISSN 2204-5538 (Online) · ISBN 978-1-925343-92-2

Demonstrating an illusory reduction in crime via sampling

James Macdonald, Data Scientist, MDSc

November 9, 2020

1 Introduction

The following is a critical review of "An evaluation of the Suspect Target Management Plan" (the report), published by the NSW Bureau of Crime Statistics and Research (BOCSAR) in October 2020. The report attempts to measure the impact of the Suspect Target Management Plan (STMP) on individual criminal behaviour. This critique will concentrate on confounding bias, and demonstrate how the reports conclusions are consistent with a nil-treatment-effect scenario.

The study hinges on an observed reduction in criminal activity after placement on an STMP compared to the year leading up to an STMP. The author has done an admirable job, working with a difficult dataset, and attempted to control for detection and selection bias by concentrating purely high-visibility crimes with a greater likelihood of reporting.

Unfortunately, there is still critical, unaddressed confounding bias, which undermines the conclusions of the study. By using people pre-STMP individuals as a control group for those post-STMP, the analysis overlooks the inherent relationship between the dependent variable, offending, and the treatment variable, STMP. The construction of the dataset introduces a confounding bias by observing individual behaviour conditioned on commencement of an STMP.

Via the below experiment, we demonstrate how an observation window that depends on the variable being measured can introduce a confounding bias into the data, and how any subsequent models will also be biased.

Concisely, this will demonstrate that for a given event E, time t and experiment commencement time T, the dataset will satisfy the inequality:

$$\frac{P(E|t < T)}{P(E|t > T)} > 1$$

if the observation point depends on such an event occuring. We will demonstrate that this inequality occurs even when individual behaviour is unchanged between target and control. Futher, we demonstrate that this will produce a bias in a subsequent models.

[1]:
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
from scipy import stats
from sklearn.linear_model import LogisticRegression
import statsmodels.api as sm
```

```
%matplotlib inline
```

1.1 Assumptions

We take 100,000 persons have individually varying criminal tendencies. We express this by articulating a threshold beyond which an individual commits a crime, drawn from a highly-skewed beta distribution and denoted by crim_thresh. We draw a similar STMP commencement threshold stmp_thresh from a uniform distribution. If a crime occurs, it lowers the probability of evading STMP by $\frac{1}{1+(n_c*f_c)}$ where n_c is the count of crimes and f_c is a severity factor.

[2]:
```
%%time
np.random.seed(1)

n_ind = 100000 # Simulate 100K individuals
p_stmp = 0.0001 # Given no crime, individual's probability of being put on a␣
↪STMP on a given day
crime_thresh = stats.beta(a=100,b=1).rvs(size=n_ind) # For each individual, set␣
↪a daily crime factor via beta dist.
stmp_thresh = stats.uniform().rvs([n_ind,1000]) # For each individual, set a␣
↪daily stmp threshold.
decay = 200 # how long it takes chance of STMP to revert
crime_mult = 2 # Factor-decrease of chances of evading STMP from crime:␣
↪Prob~(1-1/ (fctr*crime))
```

```
CPU times: user 990 ms, sys: 108 ms, total: 1.1 s
Wall time: 1.1 s
```

1.2 Experiment

We then assign everyone an activity score from a uniform distribution, and where this is greater than their individual crime threshold, this becomes a crime.

Next, we calculate an STMP evasion score, and apply a base score that someone without criminal activity receives an STMP.

Where someone's STMP evasion score falls below their STMP threshold, they are assigned an STMP. We then assign that individual an observation point based on their commencement of STMP, and remove individuals who never receive an STMP.

Figure 1 shows that this data gives us the desired result - varied criminal behaviour that is positively correlated with the probability of receiving an STMP, and a decay rate of STMP likelihood post criminal behaviour.

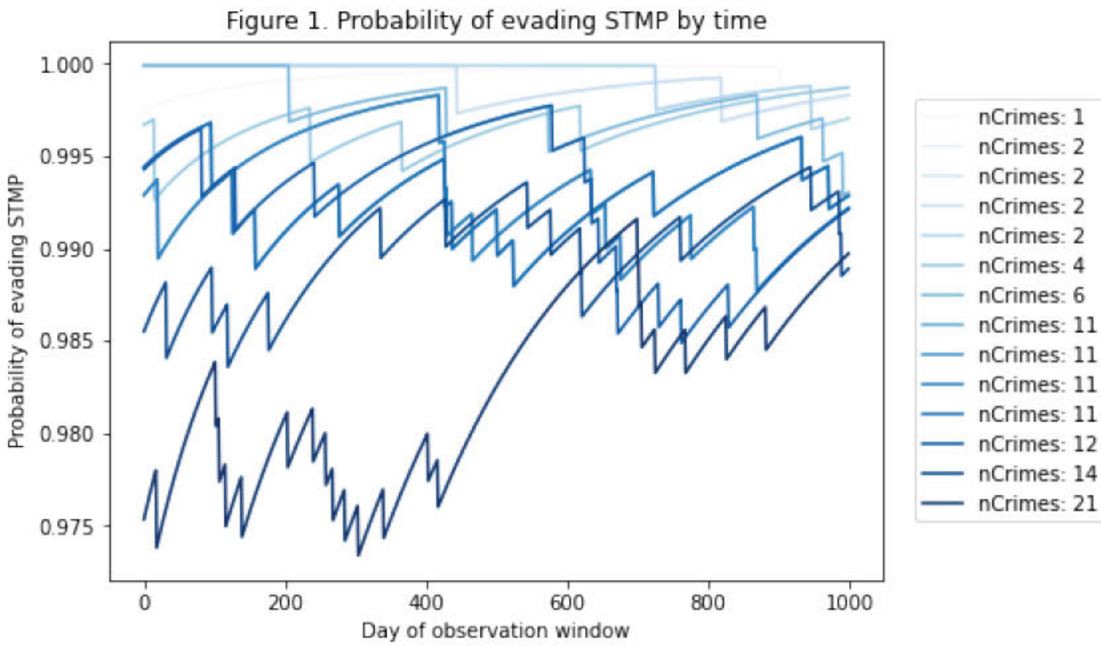
[3]:
```
%%time
activity = stats.uniform().rvs([n_ind,1000+decay]) # Roll dice on crime␣
↪activity for all individuals for 1000 days
crime = pd.DataFrame(activity.T > crime_thresh).astype(int) # crime = activity␣
↪> individual crime threshold
```

```
stmp_prop = (1/(1+crime_mult*crime) - p_stmp).ewm(halflife=decay).mean() #␣
↪probability of STMP given crime

crime.columns = [str(i) for i in crime.columns.values] # rename columns
stmp_prop = stmp_prop[decay:].reset_index(drop=True) # remove padding
crime = crime[decay:].reset_index(drop=True) # remove padding

stmps = (stmp_prop < stmp_thresh.T).astype(int) # STMP occurs where probability␣
↪> daily threshold
obs_point = stmps.idxmax() # Observation point is commencement of STMP
obs = obs_point[(obs_point>365)&(obs_point<1000-365)] # trim where stmp occurs␣
↪near edge of observation window
stmp_clean = stmps[obs.index]

observed_crime = pd.DataFrame()
for icol in stmp_clean.columns.values:
    observed_crime[icol] = crime[str(icol)][obs_point.loc[icol]-365:obs_point.
↪loc[icol]+365].values
```

```
CPU times: user 9.55 s, sys: 1.9 s, total: 11.5 s
Wall time: 11.5 s
```

[4]:
```
fig, ax = plt.subplots(figsize=(7,5))
to_plot = stmp_prop[stmp_prop.columns[:10]]
to_plot.columns = crime[crime.columns[:10]].sum().values
to_plot = to_plot[sorted(to_plot.columns)]
to_plot.columns = [f'nCrimes: {i}' for i in to_plot.columns.values]


to_plot.plot(ax=ax, cmap='Blues')
ax.set_title("Figure 1. Probability of evading STMP by time")
ax.set_xlabel("Day of observation window")
ax.set_ylabel("Probability of evading STMP")
ax.legend(loc='right', bbox_to_anchor=(1.3,0.5))
```

[4]:
```
<matplotlib.legend.Legend at 0x7f693cc696d0>
```

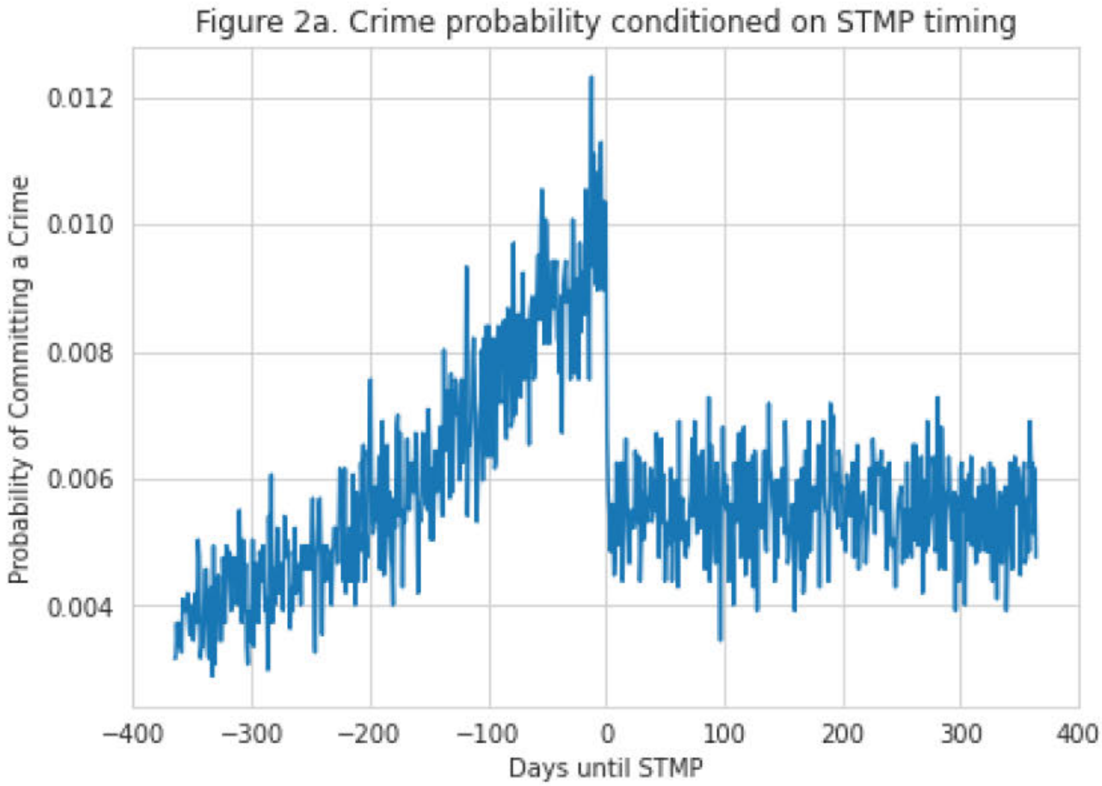


1.3 Results

We can see below that the data shows a rise in criminal activity preceding the observation point, followed by a dramatic drop. This plot is very similar to figure 1a of the report, which plots the same data.

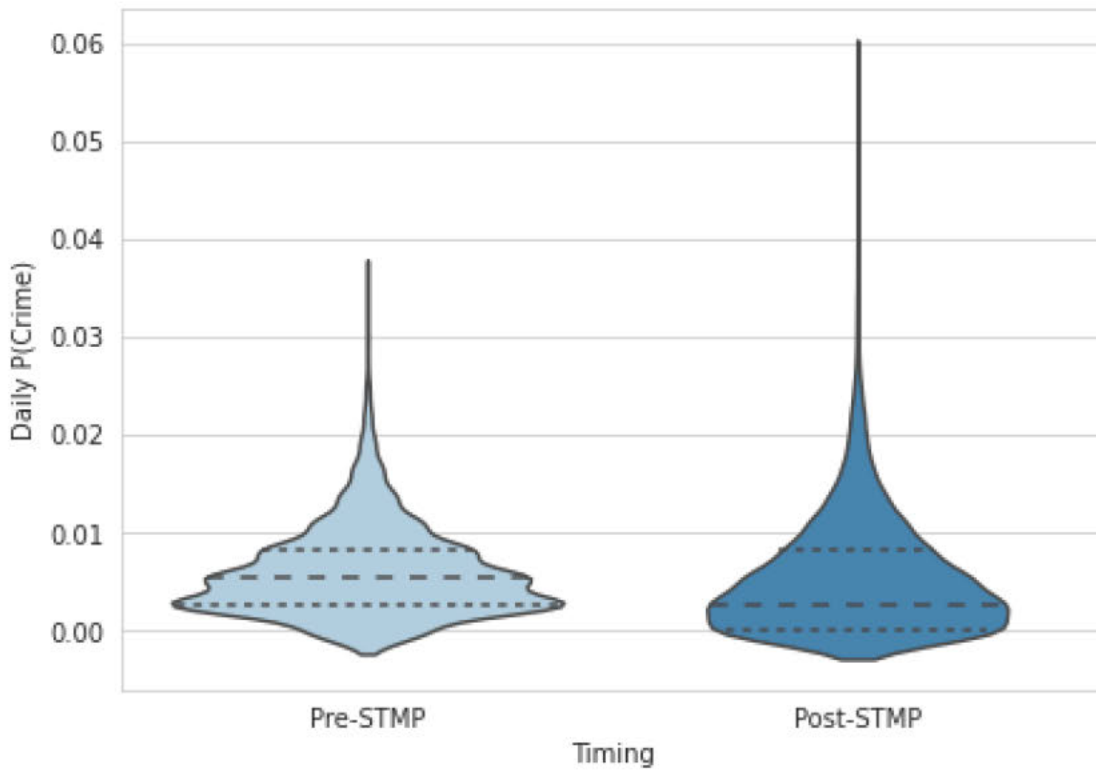
[5]:
```
sns.set_style('whitegrid')
fig, ax = plt.subplots(figsize=(7,5))
observed_crime.set_index(np.arange(-365,365)).mean(axis=1).plot(ax=ax)
ax.set_xlabel("Days until STMP")
ax.set_ylabel("Probability of Committing a Crime")
ax.set_title("Figure 2a. Crime probability conditioned on STMP timing")
```

[5]:
```
Text(0.5, 1.0, 'Figure 2a. Crime probability conditioned on STMP timing')
```




```
[6]: fig, ax = plt.subplots(figsize=(7,5))
     cols = ['Pre-STMP','Post-STMP']
     pre_and_post = pd.DataFrame([observed_crime[:365].mean(),observed_crime[365:].
      ↪mean()], index = cols).T
     for_vplot = pd.melt(
         frame= pre_and_post, value_vars = ['Pre-STMP','Post-STMP'],␣
      ↪value_name='Daily P(Crime)', var_name='Timing'
     )
     sns.violinplot(data=for_vplot, x='Timing', y='Daily P(Crime)', palette='Blues',␣
      ↪inner='quartile', bw=0.25, ax=ax)
```

```
[6]: <AxesSubplot:xlabel='Timing', ylabel='Daily P(Crime)'>
```



[7]:
```python
pre_and_post.describe()[1:].apply(lambda x: x.apply(lambda y:
    f'{round(y*100,4)}%'))
```

[7]:

	Pre-STMP	Post-STMP
mean	0.6063%	0.5531%
std	0.4579%	0.561%
min	0.0%	0.0%
25%	0.274%	0.0%
50%	0.5479%	0.274%
75%	0.8219%	0.8219%
max	3.5616%	5.7534%

[8]:
```python
def calculate_trimmed_mean(qvalue):
    pre_outliers = pre_and_post['Pre-STMP'] < pre_and_post['Pre-STMP'].
    quantile(qvalue)
    post_outliers = pre_and_post['Post-STMP'] < pre_and_post['Post-STMP'].
    quantile(qvalue)
    return pd.concat([
    pre_and_post['Pre-STMP'][pre_outliers].describe(),
    pre_and_post['Post-STMP'][post_outliers].describe()
], axis=1)
```

```
calculate_trimmed_mean(0.96)[1:].apply(lambda x: x.apply(lambda y:␣
↪f'{round(y*100,3)}%'))
```

[8]:

	Pre-STMP	Post-STMP
mean	0.541%	0.451%
std	0.353%	0.402%
min	0.0%	0.0%
25%	0.274%	0.0%
50%	0.548%	0.274%
75%	0.822%	0.822%
max	1.37%	1.37%

1.4 Interpretation

The above plots and summary statistics demonstrate that, despite zero change in underlying individual behaviour, there is a dramatic observed shift in Pre-STMP and Post-STMP criminal activity. This shift can be described as the function of the inequality between conditional probabilities, from the above table we can substitute our results and maintain the inequality:

$$\frac{P(E|t<T)}{P(E|t>T)} = \frac{\mathbf{E}[Crime_{preSTMP}]}{\mathbf{E}[Crime_{postSTMP}]} = \frac{0.61\%}{0.55\%} > 1$$

Furthermore, the median post-STMP subject has double the offending likelihood of the median pre-STMP subject. Trimmed summary statistics that exclude upper tails would further exacerbate the difference between mean values. Excluding the top 4% of outliers increases mean treatment effect from 9% to 17%.

1.5 Effect on modelling

There is no magic mechanism in modelling that can overcome this bias. Passing the data through a model will necessarily return a negative coefficient due to the relationship between the treatment and the outcome. This is demonstrated below, with a logistic regression returning a coefficient of β_{stmp} = -5.19, with a p-value significant at any level.

[9]:
```
is_stmp = pd.DataFrame(np.zeros_like(observed_crime))
is_stmp.loc[np.arange(365,730)] = 1
```

[10]:
```
X = pd.DataFrame(is_stmp.stack())
y = observed_crime.stack()
```

[11]:
```
%%time
lreg = sm.Logit(y.values, X.values).fit()
print(lreg.summary())
```

```
Optimization terminated successfully.
         Current function value: 0.363705
         Iterations 9
                           Logit Regression Results
```

```
==============================================================================
Dep. Variable:                      y   No. Observations:              7822680
Model:                          Logit   Df Residuals:                  7822679
Method:                           MLE   Df Model:                            0
Date:                Mon, 09 Nov 2020   Pseudo R-squ.:                  -9.205
Time:                        10:57:30   Log-Likelihood:            -2.8451e+06
converged:                       True   LL-Null:                   -2.7879e+05
Covariance Type:            nonrobust   LLR p-value:                       nan
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -5.1918      0.007   -761.524      0.000      -5.205      -5.178
==============================================================================
CPU times: user 41.2 s, sys: 18.4 s, total: 59.6 s
Wall time: 24.5 s
```

Furthermore, including control variables to limit confounding will not address this bias. In order to have a meaningful effect on the estimation of β_{stmp} a control variable would necessarily be correlated with the STMP variable in the data, and this has been explicitly ruled out in the report by attempting to match on proper controls. We can assume that no control variables are correlated with treatment, and thus no control variables can remove the bias in the model estimates.

1.6 Conclusion

This experiment demonstrates that an observation time triggered by a treatment variable that is correlated with the dependent variable introduces a bias into the dataset. The observed distributions of pre-treatment and post-treatment behaviour are significantly different despite, *by construction*, no change in underlying individual behaviour. It also demonstrates that the bias in the dataset will produce spurious model coefficients, and this cannot be rectified by the introduction of control variables.

In the context of the report, using pre-STMP individuals as controls for post-STMP individuals would carry the issues identified above. Furthermore, no dataset constructed from the subset of individuals on STMP is able to provide meaningful insight into the impact of STMPs, irrespective of the modelling or matching methodology used. This could be rectified by matching against individuals who are not subject to STMP in order to provide a counterfactual.

Simulation experiment for "*An evaluation of the Suspect Target Management Plan, October 2020*" Study

Dr. Gordana Popovic, UNSW, 5 November 2020

Aim

The purpose of this simulation is to see what data and analysis might look like if there was no effect of STMP on offending rates of individuals, both in terms of the plots (Figure 1a and b) and any analysis.

Executive summary

- Synthetic data was simulated in which offending did not change after entry to STMP
- Offenders in the synthetic data were recruited to STMP (a random number of days) following an offence
- Analysis from the Study was recreated with this synthetic data
- Though there was explicitly **no change** in offending following recruitment in the synthetic data, the analysis nevertheless showed strong evidence of a significant decrease in offending after recruitment, and figures consistent with the Study.
- The result in the synthetic data is an artifact of sampling, as offenders were recruited after an offence, the rate of offending just prior to recruitment is artificially increased.
- A further simulation demonstrated that even when offending **increases** post recruitment, the analysis will show strong evidence of a decrease, and figures consistent with the Study.

Method

I simulate a constant rate of offending for 10000 individuals, using an exponential distribution. The time point where individuals start STMP will be allocated a random waiting time after a randomly chosen contact with the justice system (offence), however I will not alter the offending rate of individuals in any way after they start STMP.

Some preliminaries:

- we model 10000 individuals.
- we allow each individual to have their own rate of offending, with an average over individuals of `lambda = 200` days between offences, and a standard deviation of `sig=10` between individual offending rates. Rates are kept constant over time for each individual.
- we model `n_off=20` offences to have enough to cover the two year time span, we then remove any data outside the one year before and after STMP window.
- each individual is put onto STMP an average of `wait=100` days after a randomly selected offence.

```r
library(dplyr)
library(tidyr)
library(ggplot2)
N=10000 # number of idividuals
lambda=200 #average number of days beween offences
sig=10 #individual variation in ofending rate,
# can be set to 0 for constant offending rate across individuals
```

```
n_off=20 #number of offenses modelled
wait=100 #average number of days from random offence to being put on STMP
```

Simulation

Simulating days of offences relative to entry into STMP, with no effect of STMP on offending.

```
d=matrix(NA,N,(n_off))
for(i in 1:N){
  lambda_ind=abs(rnorm(1,lambda,sig)) #individual offending rate centered around lambda
  abs=round(cumsum(rexp(n_off, rate=1/lambda_ind)),0) #day of each offense
  d_stmp<-round(sample(abs,1)+rexp(1,1/wait),0) #date of going on STMP
  rel=abs-d_stmp #day relative to STMP starting
  d[i,]=rel
}
dat=data.frame(id=1:N,d)
```

Data organisation

We then reorganize the data to long format, where there is now one column for id, one for the relative time and one for offending occasion.

```
dat_long <- dat %>%
  pivot_longer(-c(1:2), values_to= "days_before_after") %>%
  mutate(occasion=as.numeric(substr(name,2,3))) %>%
  mutate(time=ifelse(days_before_after>0,1,0)) %>%
  select(id,occasion,days_before_after,time)   %>%
  filter(days_before_after>-365) %>% #subset to 1 year before and after STMP
  filter(days_before_after<365)
dat_long
```

```
## # A tibble: 41,392 x 4
##       id occasion days_before_after  time
##    <int>    <dbl>             <dbl> <dbl>
##  1     1       14              -109     0
##  2     1       15               -27     0
##  3     1       16               300     1
##  4     2       17              -304     0
##  5     2       18              -186     0
##  6     2       19              -184     0
##  7     2       20               -11     0
##  8     3       18              -209     0
##  9     3       19                -7     0
## 10     3       20               178     1
## # ... with 41,382 more rows
```

Results

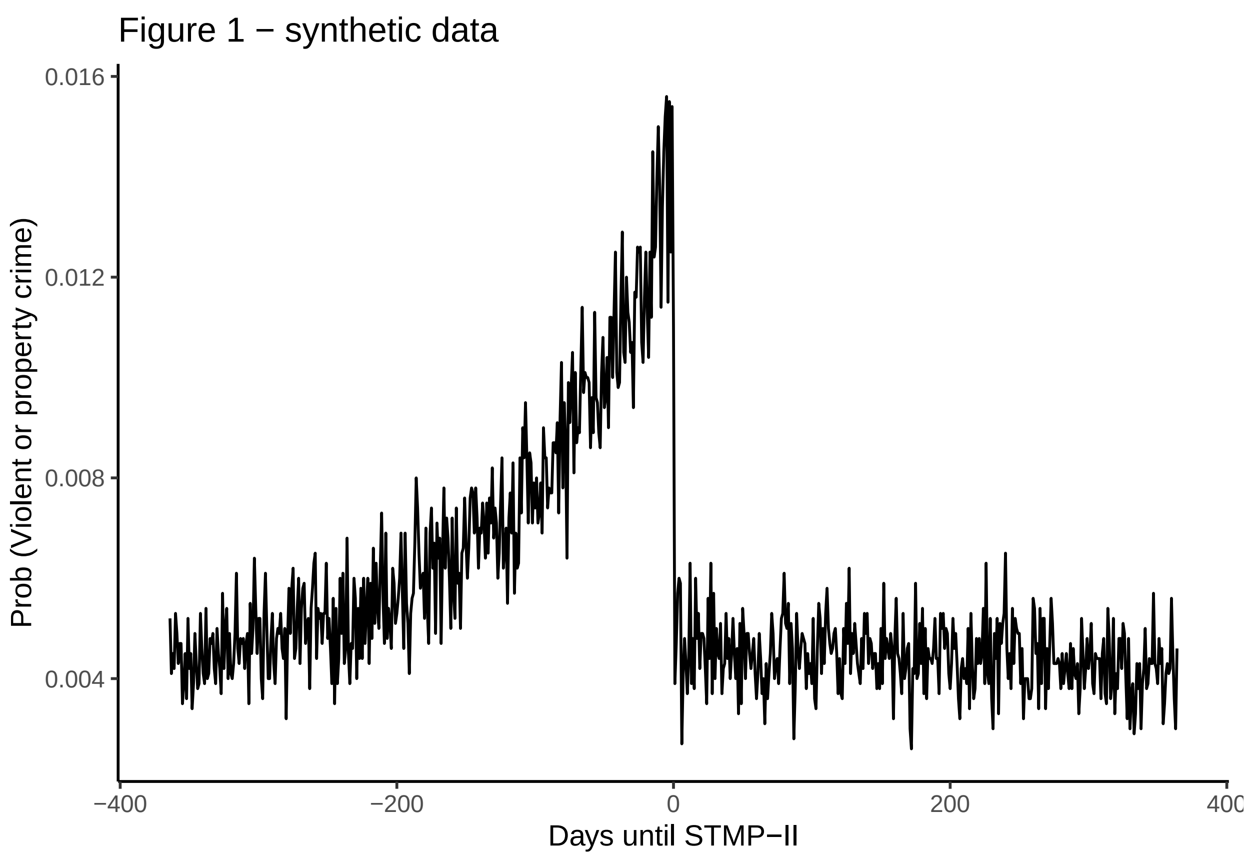
Figure 1

It is now straight forward to calculate the proportion of individuals who offended on each day.

```
crime_rates<-dat_long %>%
  group_by(days_before_after) %>%
  count()
```

And plot this.

```
crime_rates %>%
  ggplot(aes(days_before_after,n/N))+
  geom_line()+
  xlab("Days until STMP-II")+
  ylab("Prob (Violent or property crime)")+
  ggtitle("Figure 1 - synthetic data")+
  xlim(-365,365)+
  theme_classic()
```



Regression

We first calculate for each individual whether they offended in the year leading up to, and the year following them starting STMP.

```
mod_dat=dat_long %>%
  group_by(id) %>%
  summarise(before=(sum(time==0)>0)*1, #any offence in year before STMP
            after=(sum(time==1)>0)*1) %>% #any offence in year after STMP
  pivot_longer(-1, names_to = "time",values_to="offended") %>%
  mutate(time=relevel(factor(time), ref="before"))

head(mod_dat)
```

```
## # A tibble: 6 x 3
##      id time   offended
##   <int> <fct>     <dbl>
```

```
## 1     1 before        1
## 2     1 after         1
## 3     2 before        1
## 4     2 after         0
## 5     3 before        1
## 6     3 after         1
```

We then reproduce (a simplified version) of the analysis in the Study.

```r
summary(glm(offended~ id+time, data=mod_dat, family=binomial) )
```

```
##
## Call:
## glm(formula = offended ~ id + time, family = binomial, data = mod_dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6317   0.2537   0.2564   0.6994   0.7078
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.389e+00  6.884e-02  49.233   <2e-16 ***
## id           4.193e-06  7.744e-06   0.541    0.588
## timeafter   -2.133e+00  6.205e-02 -34.371   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 14939  on 19827  degrees of freedom
## Residual deviance: 13203  on 19825  degrees of freedom
## AIC: 13209
##
## Number of Fisher Scoring iterations: 6
```

We find strong evidence ($p<0.001$) of significant decrease in offending after STMP.

Conclusion

Despite simulating data with **no** effect of STMP, we see very similar results from this simulation and the Study.

- The shape of *Figure 1 - synthetic data* is very similar to Figure 1a, and particularly Figure 1b from the Study.
- There is strong evidence ($p<0.001$) of significant decrease in offending after STMP for the synthetic data, though we know that offending does not change.

We conclude that the results in the Study are consistent with a population without any effect of STMP. It seems likely the effect is an artifact of the timing of entry to STMP being after contact with the justice system.

Appendix

Simulation 2 - increased offending after STMP

This simulation aims to find a reason why in Figure 1 the rate of offending after STMP reduces over time. It seems plausible that this is because recruiting offenders to STMP temporarily increases offending, or makes them more likely to be caught, for a period following their recruitment, which thereafter goes back to their usual offending rate. The simulation below explores this scenario.

```
d=matrix(NA,N,(4*n_off))
for(i in 1:N){
  lambda_ind=abs(rnorm(1,lambda,sig)) #individual offending rate normal centered around lambda
  between_pre=rexp(n_off/2, rate=1/lambda_ind) #usual offending rate
  between_post1<-rexp(n_off/4, rate=2/(lambda_ind)) # double offending rate
  between_post2<-rexp(n_off/4, rate=1/(lambda_ind)) #usual offending rate
  abs=round(cumsum(c(between_pre,between_post1,between_post2)),0) #day of each offense
  d_stmp<-round(sum(between_pre)+rexp(1,1/wait),0) #date of going on STMP
  rel=abs-d_stmp #day relative to STMP starting
  d[i,]=rel
}
dat=data.frame(id=1:N,d)
```

And plot this.

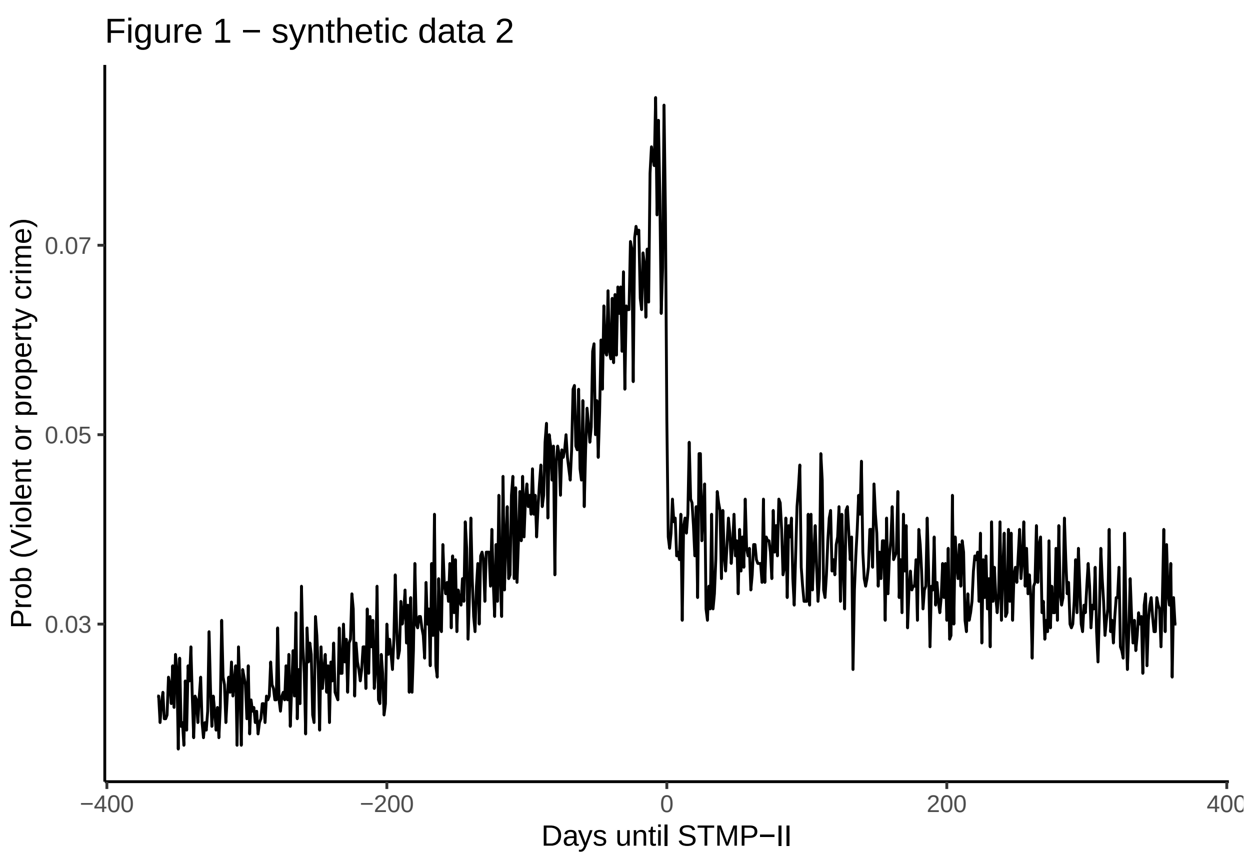


Figure 1 - synthetic data 2 has many of the properties of Figure 1a and 1b in the Study, including the apparent increase in offending prior to STMP, an apparent sharp decrease after STMP starts, and a gradual decrease in the following year. This figure was produced with synthetic data where there is no change in offending prior to STMP, then a temporary **increase** in offending after recruitment to STMP.

```
summary(glm(offended~ id+time, data=mod_dat, family=binomial) )
```

```
##
## Call:
## glm(formula = offended ~ id + time, family = binomial, data = mod_dat)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.7318   0.0467   0.0505   0.2487   0.2694
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  6.969e+00  3.191e-01  21.835   <2e-16 ***
## id          -3.142e-05  1.980e-05  -1.586    0.113
## timeafter   -3.356e+00  3.072e-01 -10.924   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3254.1  on 19919  degrees of freedom
## Residual deviance: 2903.2  on 19917  degrees of freedom
## AIC: 2909.2
##
## Number of Fisher Scoring iterations: 9
```

We still find strong evidence ($p<0.001$) of significant **decrease** in offending after STMP, even though the rate of offending has actually **increased**.

This result is consistent with the findings of the matched group analysis (page 24 of the Study) which found that *individuals subject to STMP offend at much higher rates than their matched counterparts*.

Last compiled on 05 November, 2020

A critical review of the BOSCAR report: *An evaluation of the Suspect Target Management Plan*

Ian Watson

Social Policy Research Centre

UNSW Sydney*

9 November 2020

Introduction

This paper is a critical review of a recent BOSCAR Bulletin, *An evaluation of the Suspect Target Management Plan* by Steve Yeong.[1] For ease of expression I will refer to it from now on as the *STMP Report*. My commentary covers the following issues:

- the nature of the STMP-II data;
- research design and the issue of causality;
- some technical weaknesses in the modeling.

I focus on both methodological issues as well as the interpretation of the findings, and my conclusion is that the *STMP Report* has serious weaknesses. I am particularly critical of the author's argument that his modeling shows that the SMPT-II has reduced criminality in NSW.[2] I conclude that a more accurate assessment of this study is that methodological weaknesses in the analysis have prevented any reasonable assessment being made regarding the outcomes of the STMP-II program.

*Email: mail@ianwatson.com.au. Website: ianwatson.com.au

1. Steve Yeong (2020), *An evaluation of the Suspect Target Management Plan*, Crime and Justice Bulletin Number 233, Sydney NSW: NSW Bureau of Crime Statistics and Research

2. All of my comments refer only to the STMP-II data and analysis; I do not discuss the DV-STMP data or analysis.

The nature of the STMP-II data

Cameos and typical persons

Why does the STMP-II data matter? When it comes to interpreting the findings, the nature of the data affects the reader's perception of who the STMP-II is applied to. The *STMP Report* makes it very clear who the author thinks this is:

> By the time that the typical individual is placed on either form of STMP, he has almost 10 prior court appearances, half of which relate to the use of violence, one relating to the use of weapons and two relating to the use of drugs. He has also had a sentence of imprisonment and five community orders, all by age 26 …[3]

This is essentially a cameo drawn from the sample extracted by the author from the Reoffending Database (ROD). The terminology of 'typical' would lead the reader to assume that this cameo, while not constituting the majority of the STMP-II sample, is nevertheless reasonably common. Phrased in this way, the cameo presents a disturbing and threatening picture of 'criminality' in the community. Is it an accurate account of those people subject to the STMP-II program?

The core problem here is that the report provides no information on how many people subject to STMP-II actually fit this cameo. Fortunately, it is feasible to construct *synthetic data* based on the descriptive statistics (sample size, means and standard deviations) provided in Table 1 ('First day on STMP') in the *STMP Report* for the count variables in the sample.[4] I do not argue that this synthetic data is a reconstruction of the real data; rather I argue that the distribution of possible values in the synthetic data is close to those in the real data. We don't know, for example, how these variables combine at the unit level, that is, how many individuals have a certain combination of the characteristics represented by these variables. We cannot accurately estimate, therefore, how many people are likely to fit the cameo outlined above. (There is a method, however, for simulating a unit-record dataset from these synthetic data, and I will discuss that below.)

For the moment, it is worth asking whether basing a cameo on the sample means is appropriate? In the real data, all of the count variables have standard deviations that are large relative to their means. In the synthetic data, this gives rise to distributions such as those shown in Figure 1. Even

3. Yeong 2020, p. 6.

4. The synthetic data for the count variables (all of which are overdispersed) are simulated using R's `rnegbin` function with `n` equal to sample size, `mu` (μ) equal to the mean and `theta` equal to a dispersion parameter calculated as $(\mu + \mu^2)/sd^2$.

a cursory glance shows that these data are heavily right skewed (they are overdispersed count data) with the median lower than the mean in all cases. The percentages shown on the vertical axes are illuminating: the vast majority (over 80 per cent) of the synthetic sample have zero weapons offences and a clear majority (over 60 per cent) have zero drug offences and zero prison sentences. One conclusion that can be draw from these distributions is that for individual offences (or court appearances/sentences), the STMP-II data is best characterised as: *a large number of people have a small number of offences (or court appearance/prison sentences) and a very small number of people have a large number of offences (or court appearance/ prison sentences).*

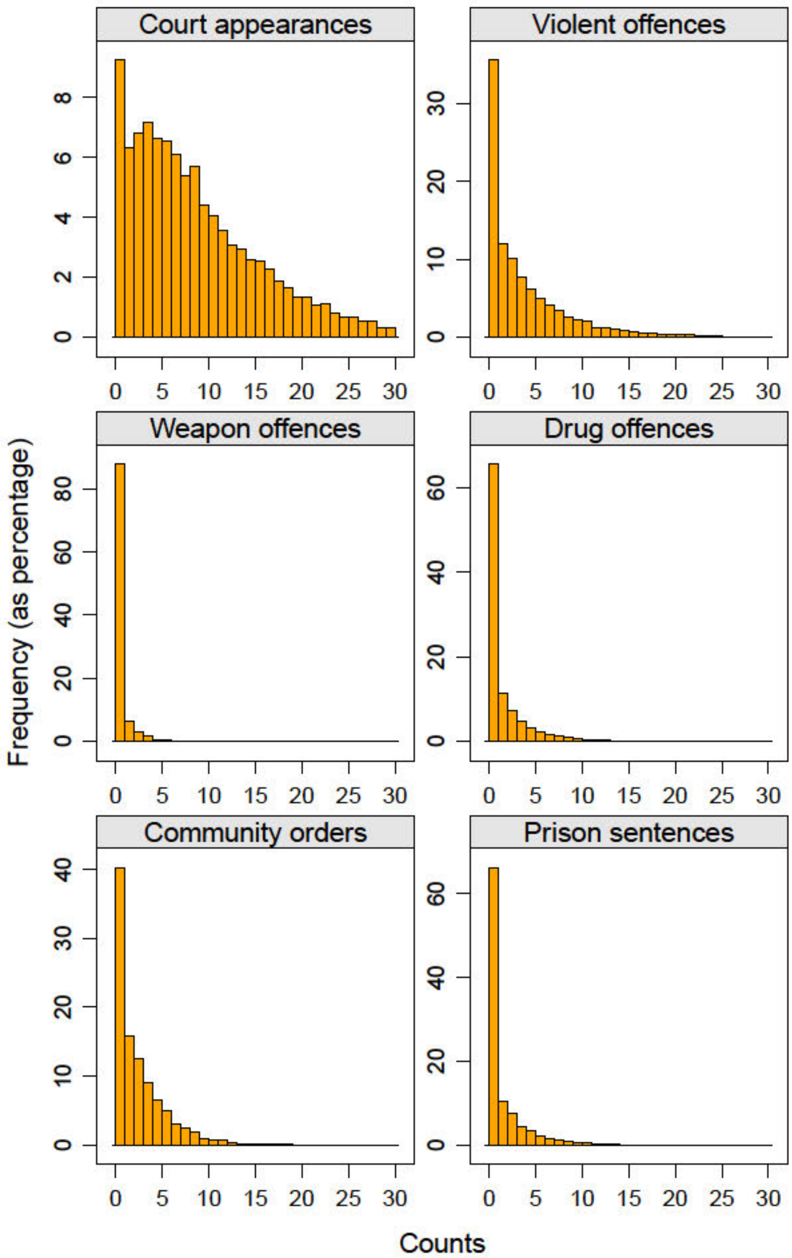


Figure 1: Distribution of count variables in synthetic data

In simulating this synthetic data—based on the descriptives shown in Table 1 of the *STMP Report*—there is little scope to change these distributions. It is certainly not possible to alter the skewed nature of the distributions. For example, 1000 iterations of the simulation for the court appearance variable consistently reproduces a strongly right skewed density. Furthermore, for these kinds of data, even if one manually alters the lower counts (replaces the 0s and 1s with 2s and 3s, for example) so as to shift the data further to the right, this increases the mean, but it reduces the standard deviation. If one attempts to retain the standard deviation by enlarging the range of high values, then this shifts the mean well above the value reported for the real data. In other words, for overdispersed count data like these there is little scope to maintain the location and scale of these variables (the mean and standard deviations) and yet to reshape this distribution away from a highly skewed shape.

The 'typical' person cameo mentioned above (referred to from now on as the '*stylised STMP-II cameo*') uses the *mean* rather than the *median*, an inappropriate measure with highly skewed data. As Table 1 shows, the means of these variables in both the original and the synthetic data differ from the median and the mode in the synthetic data. The mode is the measure which probably comes closest to the everyday notion of 'common' or 'typical' but the median is generally preferred as the most accurate reflection of the central tendency in data like these. In both cases, these figures are lower than the means, yet the means are used to construct the cameo discussed above. In other words, the severity of these interactions with the criminal justice system (CJS) is inflated in the *stylised STMP-II cameo*.[5]

Table 1: Measures of central tendency in original and synthetic data

	Original		*Synthetic data*			
Variables	*Mean*	*Rounded mean*	*Mean*	*Rounded mean*	*Median*	*Mode*
Court appearances	9.70	10	9.63	10	8	4
Violent offences	4 50	4	4.33	4	3	0
Weapon offences	0 54	1	0.52	1	0	0
Drug offences	1.66	2	1.64	2	1	0
Community orders	2.93	3	2.89	3	2	0
Prison sentences	1.64	2	1.64	2	1	0

Notes: Rounded original means are those used in the report's cameo. Except for the number of community orders and prison sentences. It is not clear where the cameo draws those figures from.

5. The cameo in the report appears to draw all its figures from Table 1 ('First day on STMP') so this has been the basis for the simulations. The number of community orders and prison sentences differ and it is unclear where these figures are drawn from.

These inflated counts are a problem, but a minor one. The more serious problem lies in the assumption in the *STMP Report* that it is reasonable to construct a typical individual from summary measures for the whole sample in an additive fashion. Constructing cameos may be intended as a device to make the descriptive statistics more vivid to a lay audience, but it can be a highly misleading device, particularly when the characteristics are combined in this additive way.

Is it possible to gain a more realistic sense of the prevalence of court appearances, criminal offences and prison sentences in the the STMP-II data rather than rely on this misleading cameo? I mentioned above that there is a method for simulating a synthetic unit record dataset from these synthetic data and, in so doing, estimate the size of the gap between what is most likely to be the case and what this *stylised STMP-II cameo* presents.

The simulation exercise proceeds as follows. Four cameos are constructed, the first of which matches the *stylised STMP-II cameo*. The other three are variations on this first one in which a more 'relaxed' definition of the combination of offences is constructed. I will say more about these shortly. The next stage of the exercise involves constructing four synthetic datasets. These reflect a number of different approaches to combining vectors (the variables) into matrices which reflect different combinations of characteristics at the unit record level. The first dataset—called the 'random dataset'—is based on repeatedly randomly shuffling the vectors so that different combinations emerge, and then counting the number of observations in the dataset for each of these four cameos. This simulation is repeated 10,000 times to produce a collection of counts, and the maximum number is then tabulated for each cameo. Why select the maximum counts? Basically, in order to favour an outcome similar to the *stylised STMP-II cameo* taking the maximum count across all 10,000 iterations makes it more likely that we will find people who combine these characteristics in a way which might approximate the *stylised STMP-II cameo*.

Of course, a random dataset like this ignores the likely correlation between these offences within individual observations. For example, individuals may be more likely to have prison sentences if they have committed violent offences. Three datasets are constructed which incorporate such correlations.[6] There is a 'low correlation' dataset in which we assume only weak correlations between all the variables. Another dataset is a 'high correlation' one, where strong correlations are assumed. Finally, a 'real

6. These datasets are constructed using copulas which preserve the marginal distributions of the variables whilst inducing correlations between them. The pairs plots in the Appendix illustrate the outcome.

world' dataset is constructed, in which correlations are differentiated in an attempt to match likely real-world conditions.[7] See the pairs plots in the Appendix for a visual representation of these four datasets.

The four cameos to which the counting exercise is applied are shown in Table 2. The first cameo matches the *stylised STMP-II cameo*. The second cameo relaxes the requirement for an exact match, by allowing any number of elements (ie. courts appearances, offences, prison etc) up to the numbers shown in the stylised cameo. The third cameo also relaxes the requirement for an exact match by allowing the elements to all be greater than those in the stylised cameo. Finally, the fourth cameo also relaxes the exact match by setting boundaries around the numbers, for example, between 8 and 12 court appearances. This last cameo is also notable in allowing for both weapons offences and prison sentences to vary from none at all (which is very common) through to two such outcomes. In other words, a range of variations on the original *stylised STMP-II cameo* are constructed in two of these cameos which favour higher counts than does this original; and one cameo (number three) which looks for 'dangerous' combinations of elements.

The results from this exercise are shown as percentages[8] in Figure 2. Despite the best efforts to match the *stylised STMP-II cameo*—and a series of alternatives—all these numbers fall way short of anything which could be regarded as 'typical'. The highest number of observations is 699 (or 7 per cent) is for cameo three—the 'dangerous' combination of characteristics—and this only applies to the dataset where all these variables are highly correlated. In other words, this number is partly an artefact of the dataset, since by construction it maximises such combinations.

In summary, despite relaxing the definition of the original *stylised STMP-II cameo* in a variety of ways, the largest proportion of people to whom it might apply is less than 7 per cent. Applying the definition as it originally appeared in the *STMP Report* sees virtually no-one fitting this cameo. By way of contrast, if one creates a cameo for someone with a handful of court appearances and just *one* other offence (or prison sentence or a few community orders), then one finds a match for 24 per cent of the 'real-world' synthetic dataset.[9] In other words, offenders with only

7. These draw on information from experts in criminology and from data in Patrizia Poletti et al. (2010), 'Common offences in the NSW higher courts', in: *Judicial Commission of NSW: Sentencing Trends & Issues*, URL: https://www.judcom.nsw.gov.au/wp-content/uploads/2016/07/sentencing_trends_41.pdf and Georgia Brignell et al. (2010), 'Common offences in the NSW local court', in: *Judicial Commission of NSW: Sentencing Trends & Issues*, URL: https://www.judcom.nsw.gov.au/wp-content/uploads/2016/08/sentencing_trends_40.pdf.

8. Since these are based on 10,000 observations, conversion into counts is simple: multiply the percentage shown by 100.

9. This cameo is not shown in Figure 2 or Table 2 but consists of 2380 observations. The

a single offence make up one quarter of this synthetic STMP-II dataset. Clearly, these simulations reinforce the view expressed earlier, that the people subject to STMP-II consist of *a small group of people with a large number (or range) of interactions with the criminal justice system (CJS) and a large group of people with a small number (or range) of interactions with the CJS*.

Table 2: Definitions of cameos

Category	*Definition*
Cameo 1	court==10 & viol==5 & weap==1 & drugs==2 & comm==5 & pris==1
Cameo 2	court % n% 1:10 & viol %in% 1:5 & weap==1 & drugs % n% 1:2 & comm %in% 1:5 & pris==1
Cameo 3	court > 10 & viol > 5 & weap > 1 & drugs > 2 & comm > 5 & pris > 1
Cameo 4	court % n% 8:12 & viol %in% 3:7 & weap % n% 0:2 & drugs %in% 0:3 & comm %in% 3:7 & pris % n% 0:2

Notes: Abbreviations: == equal to; %in% n the range; 1:10 1 to 10; > greater than. Note that Cameo 1 matches the *stylised STMP-II cameo*.

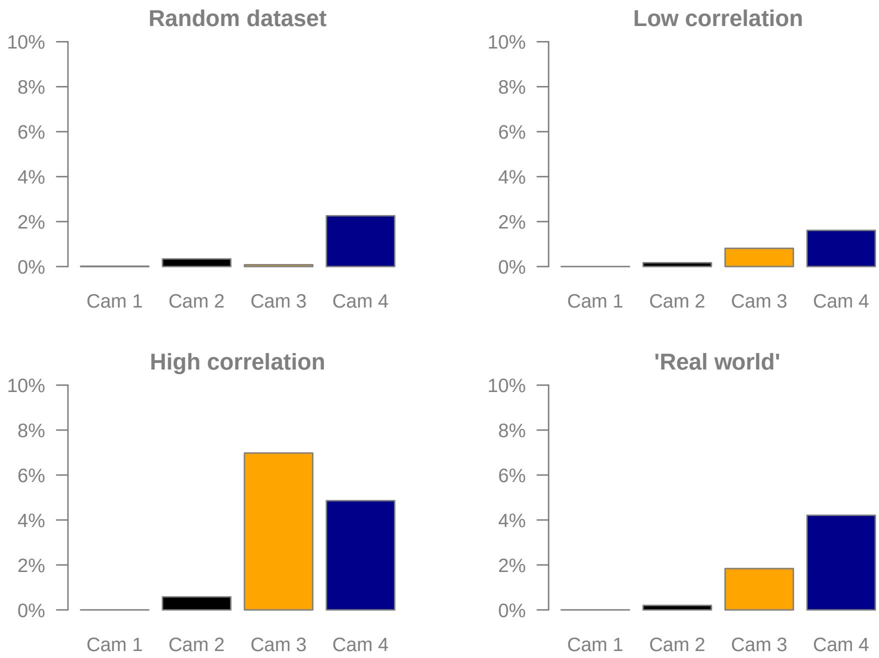


Figure 2: Percentage of observations in each dataset which match cameos
Definitions of cameos are shown in Table 2

The report argues that the descriptive statistics in Table 1 show that 'the police are identifying high-risk individuals for STMP'.[10] As suggested earlier, this claim mistakenly extrapolates from overall sample averages to construct a 'typical' individual, and this is then used to argue that a large number of people have been legitimately placed on STMP-II. Clearly, the synthetic data suggests such a claim is completely unwarranted and that the

'handful' of court appearances are for 1 to 4 and the 'few' community orders are for 1 to 3.
10. Yeong 2020, p. 6.

STMP-II program is wide-ranging in its application rather than precisely targeted.

The onus lies with the author of the *STMP Report* to refute this conclusion by *using the real data* to generate counts for the cameo in that report and to display the distributions shown in the real data along the lines of Figure 1 above. In other words, it rests with the author of the *STMP Report* to show that the cameo of the typical person placed on the STMP constitutes more than a small handful of people. Otherwise, one can only conclude that the *stylised STMP-II cameo* is a complete fiction.

Research design and causality

The *STMP Report* is located within the treatment effects tradition, in which a a treatment group (participants in a program) is exposed to a treatment to which a control group (non-participants) is not exposed, and one then compares outcomes across the two groups. While this approach can be applied reasonably well within an experimental setting, for observational data this approach can be fraught with difficulties, particularly when regression modeling is solely relied upon for establishing causality.[11] The usual procedure is to include a dummy variable (treated or not treated) and test whether it has a significant association with the outcome (such as offending). A range of confounding variables are also included in order to isolate the 'effect' of treatment on participants.

Research design

Counterfactuals are fundamental to assessing treatment effects within observational studies. They address the obvious question: what would have happened in the absence of treatment? It is the counterfactual which confers 'causality' on the research findings.[12] For a counterfactual to have validity the control group must be comparable on a range of variables, with the only notable difference being exposure to the treatment. The author of the *STMP Report* recognises at the outset selection bias makes it difficult to construct a valid comparison, because 'individuals on STMP are likely to be at a higher risk of offending, irrespective of whether STMP has any impact on offending'.[13] The *STMP Report* takes two approaches to this

11. Paul R. Rosenbaum (2002), *Observational Studies*, New York: Springer.

12. As the Neyman-Rubin causal model puts it: 'A causal effect is defined as the difference between an observed outcomes and its counterfactual.'Alexis Diamond and Jasjeet S. Sekhon (2013), 'Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies', in: *Review of Economics and Statistics* Vol. 95. No. 3, pp. 932–945, URL: http://sekhon.berkeley.edu/papers/GenMatch.pdf, p. 4

13. Yeong 2020, p. 7.

problem and while findings for each are presented in the report, the second approach is relegated to the appendix. I will return to this issue below.

In the first approach the study does not employ a 'conventional' control group. Rather the research design involves a time-shift strategy in order to create a control group. This group consists of another group of individuals who will be *subsequently* placed on the STMP-II but who, during the 'observation period', are not yet subject to that program.[14] However, because the treatment variable is causally dependent on the dependent variable, pre-STMP individuals cannot serve as a control group for a post-STMP target group. Because of the time-shift imposed by the study design, the number of court appearances and offences etc are *rising* among the control group while they are *falling* among the treatment group. This artefact of the study design will inevitably bias the regression towards finding a larger gap between the two groups than might otherwise be the case. This problem is not a minor one but is inherent to the research design because of the construction of the control group. Since prior offending is listed as a trigger for an STMP in the report, this study only shows that offences cause STMPs.

The author is aware of this problem. The dummy variable in the regression modeling which represents placement on STMP-II is identified through variation in the timing of when individuals become subject to SMTP-II. To interpret this dummy variable as causal relies on the risk of offending being unconditionally related to the timing, something which is not the case with these data.[15] As the author concedes, the risk of offending by individuals is not time-invariant, but rather appears to be relate to when such individuals are placed on STMP-II. The author of the *STMP Report* acknowledges this weakness in the research design:

> If [STMP] were to have a causal interpretation, we would expect so see no trend in offending prior to STMP, followed by a sharp (downward) trend after placement on STMP … However, from Figures 1a and 1b we can see sizable upward trends in the year leading up to STMP, followed by sharp downward trends immediately after being placed on STMP.[16]

The second approach—the one relegated to the appendix—entails using a matching estimators strategy to explicitly create a control group who are *not* subject to the STMP-II. These consist of a group drawn from the Reoffending Database (ROD) but who have *not been placed on STMP-II at all* (as opposed to a group placed on STMP-II in the next time period as

14. That is, the time shift involves multiple periods with pre-treatment and post-treatment groups aligned.
15. Yeong 2020, p. 8.
16. Ibid., p. 8.

happens in the time-shift strategy). Research designs based on matching estimators are well established in the literature, and as long as the researcher achieves good balance on the covariates between treatment and control groups, then regression modeling may proceed with reasonable confidence.

In implementing the matching estimators, the study combined Coarsened Exact Matching (CEM) and Propensity Score Matching (PSM), but the author was dissatisfied with the matching results: 'the matched groups are not statistically or practically equivalent to their respective treatment groups'.[17] It is not clear what 'practically equivalent' means. Perhaps the author is referring to an earlier footnote where he observed:

> Interestingly, I was not able to find a credible match for individuals subject to STMP using the entire Reoffending Database (which contains information for every person charged by the NSW Police Force since 1996). This suggests that the people the police select for STMP are truly distinct from other offenders they interact with.[18]

However, it is also likely that the matching strategy employed by the author was inadequate. I return to this issue below. The lack of a 'statistically equivalent' match is not explained in any detail. The descriptive comparison of treatment and control groups in Table A1 of the appendix does not provide compelling evidence that the two groups are not reasonably comparable. The means are shown to three decimal points, whereas if they were shown to one decimal point, the impression of how well they matched might be quite different. For example, age differences (26.7 to 26.3) equate to a few months apart, and differences for prior court appearances (10.6 to 10.1), prior prison sentences (1.8 to 1.7) and prior community orders (3.2 to 3.1) are all fairly minor. Moreover, with over 9,000 observations in each group, minor differences are almost bound to be 'statistically significant'. What matters with the matching estimator approach is whether the differences between a treatment group and a control group substantively shrink during the matching process such that one is ultimately comparing 'like with like' across a large majority of the variables employed.

It is more likely that the author's matching strategy has let him down. It is well known that propensity score approaches can worsen the matching outcome, and recent literature has reiterated this criticism.[19] Far better matching estimators are available which the author might have employed, such as 'genetic matching', an approach which invariably improves

17. Yeong 2020, p. 23.
18. Ibid., fn. 21, p. 7.
19. See, for example, Gary King and Richard Nielsen (2019), 'Why Propensity Scores Should Not Be Used for Matching', in: *Political Analysis* Vol. 27. No. 4, pp. 1–20

on propensity score outcomes.[20] In other words, instead of giving up on the matching estimator approach—and relegating the findings to the appendix—the author should have persisted with this strategy.

What is particularly disturbing is that the regressions fit to these matched data provided results opposite to those in the main body of the report. It showed higher offending among the STMP-II group compared with their counterparts. Rather than view these results as casting doubt on the main findings in the report, the author speculated that the weaknesses in the matching process may be the reason: 'One explanation for this finding is that there is some form of unobserved heterogeneity that matching cannot address'.[21] While it can be difficult to make matching estimators work well, the finding here is *not* that there is *no difference* between the groups, but that the results are the *reverse* of the findings in the main report. This anomaly surely warranted further investigation rather than a curt dismissal of the matching estimator procedure, particularly when better approaches were available.

Causality

The interpretation of the study's finding is one of the most worrying aspects of the *STMP Report*. Having concluded that he had failed in his efforts to construct a valid counterfactual, the author concluded: 'my estimates do not have a causal interpretation. Instead, they must be interpreted as the association between STMP and offending'.[22] However, a second conclusion immediately contradicted this:

> And second, this would suggest that the police are both correctly identifying individuals at a high risk of offending for STMP, and that once placed on STMP, an individual's risk of offending drops dramatically.[23]

The wording of this last sentence is clearly a causal one. This is not an isolated lapse in expression. The author repeats the caveat about causality in the discussion section of the report ('the estimates do not have a causal interpretation') but again negates this by discussing the possible direction

20. See, for example, Jasjeet S. Sekhon (2011), 'Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R', in: *Journal of Statistical Software* Vol. 42. No. 7, pp. 1–52, URL: http://www.jstatsoft.org/v42/i07/. A recent Productivity Commission report on the youth labour market made extensive use of this approach. See Catherine de Fontenay et al. (2020), *Climbing the jobs ladder slower: Young people in a weak labour market*, Staff Working Paper, July, Productivity Commission.
21. Yeong 2020, p. 24.
22. Ibid., p. 8.
23. Ibid., p. 8.

of bias in these estimates by referring to the 'true crime reduction benefit associated with STMP'. He concludes the paragraph with 'It is, therefore, likely that STMP-II is reducing other types of crime in addition to those examined in this paper'.[24] In the report's overview, the Results section is careful to stick with the language of 'association' but the Conclusion section immediately overturns this: 'Both STMP-II and DV-STMP are effective in reducing crime. Both programs predominately reduce crime through deterrence'.[25] Clearly, causal language is endemic to the author's interpretation of his results.

Even were the author to avoid the language of causality, and stay strictly with the language of 'association', the conclusion that the STMP-II had a positive and sizable association with a reduction crime is completely unfounded. There are several reasons for this:

◁ the strength of any association is indeterminate because of the bias in the time-shift comparison between treatment and control group (as outlined above);

◁ the association is the opposite in the matching estimators approach, and no serious engagement with these results is offered;

◁ the specific regression findings are quite diverse, but a *single* policy conclusion is drawn.

This last point is an important one. Using the time-shift strategy, the author finds mixed results. The associations between STMP-II and subsequent offending are:

◁ negative for property crime by the whole sample;

◁ indeterminate for violent crime by the whole sample;

◁ positive for imprisonment by the whole sample;

◁ negative for both violent and property crime for juveniles;

◁ positive for violent crime for Aboriginal participants;

◁ negative for property crime for Aboriginal participants;

When it came to the matching estimators strategy, as just noted, the associations were positive for violent and property crime (combined). In other words, negative associations (that is, a 'reduction' in crime) was far from universal, yet the report's main conclusions ignore this unevenness in the results and assert confidently that STMP-II is 'effective in reducing

24. Yeong 2020, p. 17. This constant lapsing into causal language is found throughout the report, sometimes in the context of discussing technical points. For example, the author suggests on page 8 that the estimates in the study might be conservative and underestimate 'the true impact of the STMP's crime reduction benefit.' 'Reduction' is clearly a causal term. The author then proceeds to ask: 'If STMP is generating a reduction in crime, the question is how?' The terminology of 'generating' is thoroughly causal.
25. Ibid., p. 1.

crime'. In other words, the policy implications of the report amount to an endorsement of the STMP-II, yet the regression modeling fails to support this blanket conclusion. The most accurate conclusion to this report would be: *methodological weaknesses in the analysis have prevented any reasonable assessment being made regarding the outcomes of the STMP-II program*.

It is worth noting that one of the key insights in the author's overview of the literature is that overseas programs which target subpopulations in an effort to reduce crime often supplement the policing strategy with increased social support (eg. housing, education, employment) for those subpopulations. The author cites the example of the community Initiative to Reduce Violence (CIRV) in Glasgow as one successful program that has been 'rigorously evaluated'.[26] However, towards the end of the *STMP Report* when discussing the importance of his findings, the author ignores this insight and argues:

> The first is to illustrate that offender-focused policing programs work in Australia. This is an interesting finding given that STMP differs markedly from most focused deterrence programs overseas. Focused deterrence programs typically involve working with community organisations to communicate an explicit message of deterrence. Focused deterrence programs also generally involve increasing access to social services as an adjunct to intensive policing.

In claiming that his study has shown that the STMP-II has 'worked' and that it 'caused' crime to fall the author dispenses with the relevance or need for social support in addition to policing activity. His report can be seen as an endorsement of a policing-only approach, even though it is clear that the study has not established such a causal link.

Technical weaknesses

Model fit

How well do these models fit the data? The author offers very little information on model diagnostics. The adjusted R-squared figures are nearly all below 0.1, which means that some 90 percent of the variability in the outcome is not accounted for by the predictors used in these models. A great deal else is going on in these data that is not captured well in this modeling.[27]

Other measures of fit, in particular, predictive adequacy (for example, cross-validation) are not canvassed. To some extent, the author addresses

26. Ibid., p. 4.

27. It is interesting to note that the adjust R-squared figures are higher for the regressions of the matched estimators approach compared with the regressions in time-shift approach.

this in the footnote which contrasts the objectives of prediction versus causal inference.[28] His objective is the latter and this focus can be used to justify a lack of concern with the predictive accuracy of the modeling. However this does not mean that a poor fitting model is acceptable. As Hilbe cautions:

> The problem is that predictor p-values may all be under 0.05, or may even all be displayed as 0.000, and yet the model can nevertheless be inappropriate for the data. A model that has not undergone an analysis of fit is, statistically speaking, useless.[29]

Heterogeneity

The concept of heterogeneity—diversity—is an important one in statistics and its relevance has been increasing in recent years.[30] When it comes to the treatment effects literature—the field in which the *STMP Report* can be located—there is an increasing recognition that the average treatment effect (ATE) of some intervention on the treated is not necessarily very useful. A more interesting question is: for whom did the treatment work? and for whom didn't it work? And why?

This suggests that focusing on heterogeneity should be a major focus whenever the subjects in a dataset show diversity. It is clear from the discussion above that the people subject to STMP-II are indeed quite heterogeneous. The *STMP Report* report does acknowledge heterogeneity and it does this by running separate regressions for young people and for Aboriginal people and separate regressions for a number of cohorts (based on the duration of their sentences). Unfortunately, the author's implementation of this may be unsound: he compares coefficients from separate regressions, a procedure whose validity depends on assumptions about the sample variances. A more rigorous way to deal with heterogeneity is to fit a *single* model and use either interaction terms or a multilevel model. In this way, one can answer the question: how does the relationship between outcomes and predictors vary across subgroups? In so doing, it is legitimate to make direct comparisons because all the coefficients (or predictions) come from the same model.

Another source of heterogeneity in this study are the Police Area Commands (PAC). Not only does selection into the STMP-II depend on

28. Yeong 2020, p. 10, fn 25.
29. Joseph M. Hilbe (2011), *Negative Binomial Regression*, Second edition, Cambridge: Cambridge University Press, p. 64.
30. See the emphasis on moving beyond averages in the field of quantile regression (Roger Koenker (2005), *Quantile Regression*, New York: Cambridge University Press) or the emphasis on multilevel models for investigating heterogeneity (Gelman and Hill 2007.

decisions made at the PAC level, but the subsequent interactions between the PAC and these people would appear to be quite fundamental. The modeling in the *STMP Report* regards the PAC as a 'control', specifically as a fixed effect. But is this an adequate way to deal with such heterogeneity? The characteristics of the PACs are extremely diverse, given their geographical basis. The *STMP Report* certainly recognises that heterogeneity arises from 'PAC-specific considerations such as their priority crimes, annual budgeting allocations, variation in the application of STMP-II, local labour market conditions and the demographic characteristics of civilians living within the jurisdiction of each PAC'.

The author deals with this by including the PAC as a fixed effect, a statistical device for adjusting for this variability in so far as the outcome is concerned. For example, do the characteristics of the PAC relate to the outcomes such as committing a violent or property crime or being imprisoned (see Table 2 in the *STMP Report*). But the variability in the predictors are ignored with this approach. Fixed effects cannot answer questions such as: how do the variability in age, Aboriginality, cautions, court appearances, PAC and so forth *interrelate*? How do these different covariates operate for different subgroups within the model? In other words, many of the various subgroup effects for the key predictors are not canvassed in these regressions.

To achieve this one needs interactions in a model. However, introducing the PACs as fixed effect interaction terms in not feasible—given how many PACs there are—so the obvious solution is a multilevel model in which the PAC is a grouping term (or level). From the model equation ($y_{ipt} = \beta_0 + \beta_1 Post_{ipt} + \gamma X'_{it} + \lambda_{pt} + u_{ipt}$) and the accompanying description, it is clear that the data are already indexed by PAC, so using multilevel models to accommodate this hierarchical structure is completely feasible. It is also evident that there is clustering in the sample: observations drawn from the same PAC in the sample will have greater similarity to each other than to those in other PACs. This can violate the regression assumption regarding independent error terms. The *STMP Report* acknowledges the clustering for the PAC variables and presents robust standard errors to deal with this. This approach, while adjusting the naïve standard errors, leaves the coefficient estimates unchanged. By contrast, multilevel models not only adjust the standard errors, but also improve the accuracy of the coefficient estimates.[31] In other words, not only would multilevel modeling

31. The increased accuracy comes from the 'partial pooling' which multilevel models employ. By contrast, the classical regression model, as employed in the *STMP Report* is essentially a 'complete pooling' model. For further elaboration on this distinction see ibid. One view of a multilevel model is that it operates as a 'giant interaction machine' (Richard McElreath (2020), *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, Second Edition, Boca Raton: CRC Press, Taylor & Francis Group.

offer insights into the heterogeneity in these data, but such an approach would provide better estimates.

As mentioned earlier, the author's main approach to this heterogeneity is separate regressions for subgroups. Yet when different model results are found for one of these subgroups—Aboriginal people—-the author gains little insight from his modeling and instead resorts to speculation which has no grounding in the data itself: 'Aboriginal people may react negatively to STMP-II interactions with police which results in increased offending.'[32]

In summary, there is insufficient material in the *STMP Report* to assess the adequacy of the modeling. The appendix is more of a supplement than a compendium of detailed model results and it is unclear what diagnostics the author used to assess the models. While his use of classical regression models (OLS) is standard practice in econometrics, among statisticians there is an increasing recognition that better model estimates come from using multilevel models.

References

Brignell, Georgia, Zeinab Baghizadeh, and Patrizia Poletti (2010), 'Common offences in the NSW local court', in: *Judicial Commission of NSW: Sentencing Trends & Issues*, URL: https://www.judcom.nsw.gov.au/wp-content/uploads/2016/08/sentencing_trends_40.pdf.

Diamond, Alexis and Jasjeet S. Sekhon (2013), 'Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies', in: *Review of Economics and Statistics* Vol. 95. No. 3, pp. 932–945, URL: http://sekhon.berkeley.edu/papers/GenMatch.pdf.

Fontenay, Catherine de, Bryn Lampe, Jessica Nugent, and Patrick Jomin (2020), *Climbing the jobs ladder slower: Young people in a weak labour market*, Staff Working Paper, July, Productivity Commission.

Gelman, Andrew and Jennifer Hill (2007), *Data analysis using regression and multilevel / hierarchical models*, Cambridge: Cambridge University Press.

Hilbe, Joseph M. (2011), *Negative Binomial Regression*, Second edition, Cambridge: Cambridge University Press.

King, Gary and Richard Nielsen (2019), 'Why Propensity Scores Should Not Be Used for Matching', in: *Political Analysis* Vol. 27. No. 4, pp. 1–20.

32. Yeong 2020, p. 17.

Koenker, Roger (2005), *Quantile Regression*, New York: Cambridge University Press.

McElreath, Richard (2020), *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, Second Edition, Boca Raton: CRC Press, Taylor & Francis Group.

Poletti, Patrizia, Zeinab Baghizadeh, and Pierrette Mizzi (2010), 'Common offences in the NSW higher courts', in: *Judicial Commission of NSW: Sentencing Trends & Issues*, URL: https://www.judcom.nsw.gov.au/wp-content/uploads/2016/07/sentencing_trends_41.pdf.

Rosenbaum, Paul R. (2002), *Observational Studies*, New York: Springer.

Sekhon, Jasjeet S. (2011), 'Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R', in: *Journal of Statistical Software* Vol. 42. No. 7, pp. 1–52, URL: http://www.jstatsoft.org/v42/i07/.

Yeong, Steve (2020), *An evaluation of the Suspect Target Management Plan*, Crime and Justice Bulletin Number 233, Sydney NSW: NSW Bureau of Crime Statistics and Research.

Appendix

The figures on the following pages show pairs plots for the synthetic datasets. Correlations are shown numerically in the upper triangles and visually as regression lines fit to scatter points in the lower triangles. The distributions of each variable are shown along the diagonal.

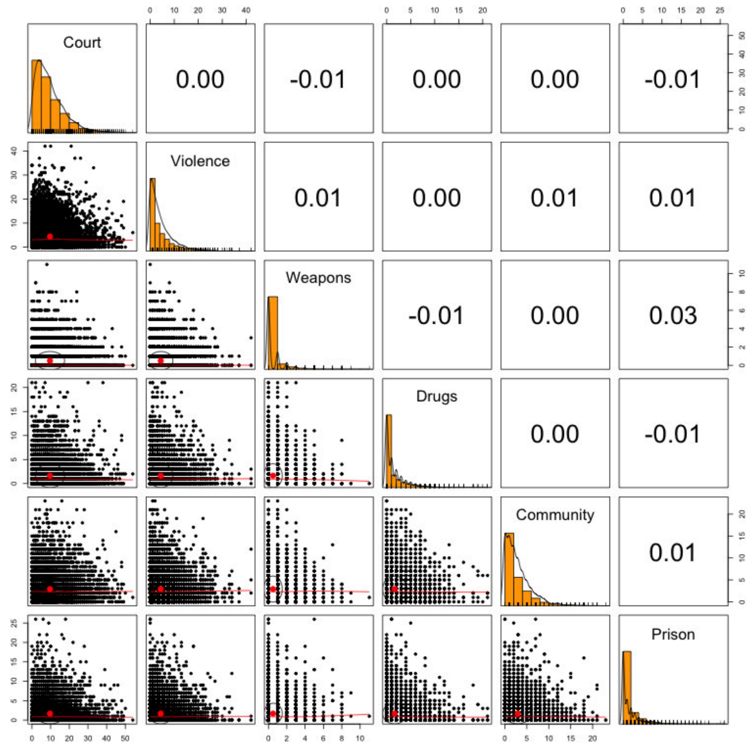


Figure 3: Pair plots of random dataset

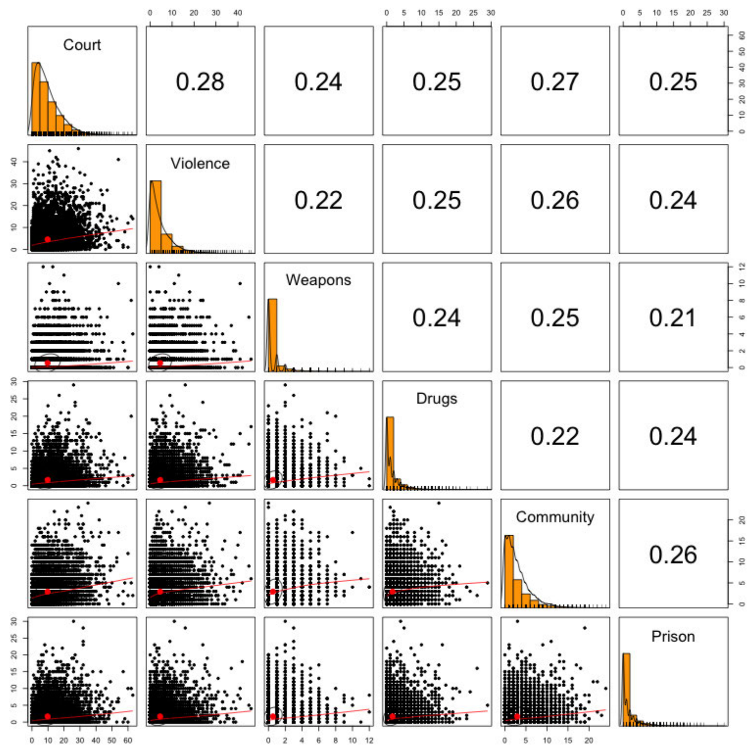


Figure 4: Pair plots of low correlation dataset

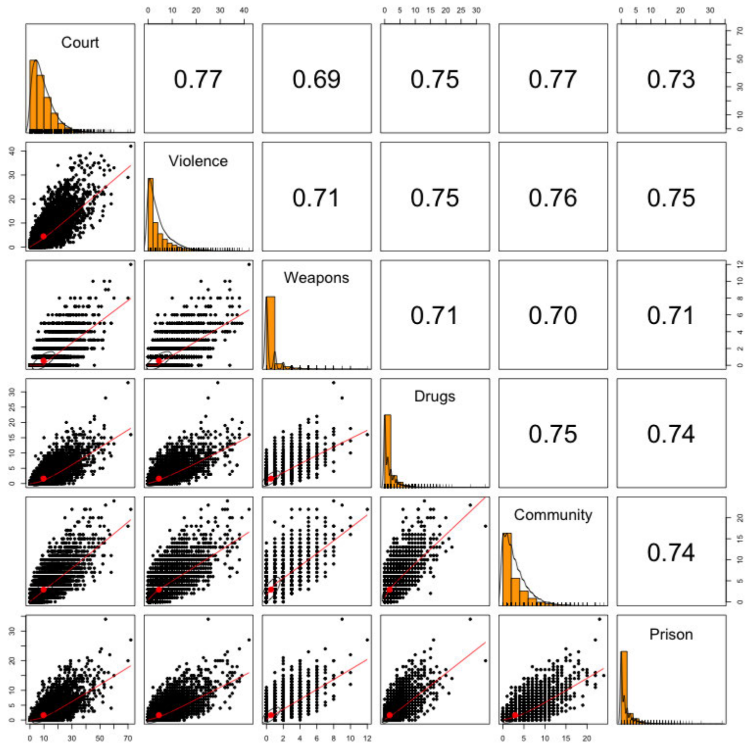


Figure 5: Pair plots of high correlation dataset

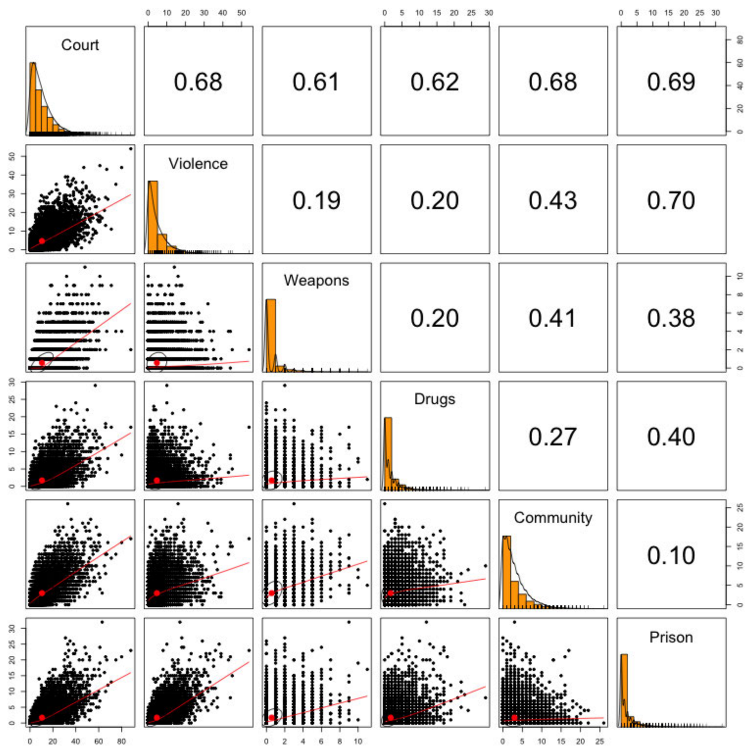


Figure 6: Pair plots of 'real-world' dataset

Response to comments on 'An evaluation of the Suspect Target Management Plan - Crime Justice Bulletin 233'

NSW Bureau of Crime Statistics and Research

February 2021

Introduction

In October 2020, the New South Wales (NSW) Bureau of Crime Statistics and Research (BOCSAR) published a Crime and Justice Bulletin titled *'An evaluation of the Suspect Target Management Plan (STMP)'*, hereafter referred to as Yeong (2020). This bulletin reported the results from the first comprehensive study of the STMP program as it currently operates in NSW. The report found that placement on STMP was associated with large, statistically significant reductions in crime. Since then, three papers (Macdonald, 2020; Popovic, 2020; Watson, 2020) critical of different aspects of the report, have been received by BOCSAR. This document responds to the claims made by these authors.

In our view, the criticisms of the STMP report raised by Macdonald (2020), Popovic (2020) and Watson (2020) can be divided into five categories:

1. Issues associated with the primary model.
2. Matching.
3. Descriptive statistics.
4. Causal claims.
5. Other technical issues.

The remainder of this response will address each of these criticisms in turn.

Issues associated with the primary model

All three papers argue that because placement on STMP is an increasing function of offending, comparing the offending behaviour of individuals in the 12 months before and after placement on STMP means that Yeong (2020) would have necessarily found a negative association between STMP and crime.

In the economics literature, this problem is known as Ashenfelter's Dip (Ashenfelter, 1978; Ashenfelter and Card, 1985). Ashenfelter's Dip originated in the job training literature, where the objective is typically to identify the causal effect of a job training program on future earnings. In this context, it was observed that participants often experienced a (possibly) transient reduction in earnings just prior to participating in a training program. If the reduction was transient, then any subsequent increase in earnings after the program could simply reflect regression to the mean. Said differently, the pre-program dip in earnings would have been followed by an increase in earnings, irrespective of participation in the program. The implication is that this issue can cause a researcher to overestimate the impact of a given program.

In our case the dip is in fact a hump (i.e., instead of a potentially transient reduction in earnings, we have a potentially transient increase in offending). We agree that this is a limitation of the identification strategy which was not considered in the original version of the paper. We also agree that this problem possibly resulted in Yeong (2020) reporting estimates that overstated the negative association between STMP and crime.

However, there are two reasons why we do not accept that the identification strategy is so biased such that we would necessarily have found a negative association between STMP and crime. First, Yeong (2020) included an important set of control variables related to both selection for STMP and crime; and second, once we exclude the hump from the estimation sample, we still find a negative association between STMP and crime. The remainder of this section provides more detail with respect to each of these points.

As Macdonald (2020) points out on page 8 of his review:

> *'In order to have a meaningful effect on the estimation of β_{stmp} a control variable would necessarily be correlated with the STMP variable in the data, and this has been explicitly ruled out in the report by attempting to match on proper controls.'*

It is true that Yeong (2020) was unable to identify a suitable set of individuals not subject to STMP using his matching strategy. However, this does not mean that the set of control variables included in his regressions are uncorrelated with placement on STMP. The control variables used by Yeong (2020) were explicitly chosen to act as proxies for the (observable subset of) factors that police consider when selecting individuals for STMP.[1] The utility of these control variables is evident in Table 2 of Yeong (2020). For example, in Panel A of Table 2, we can see that inclusion of the control variables reduces the absolute size of the point estimate by 2.2 percentage points (26.5% in relative terms).

In order to provide some empirical evidence as to whether the entire association between STMP and crime reported in Yeong (2020) is driven by the hump, we follow a similar approach to Machin and Marie (2011) and employ five robustness checks:

1. Examine offending within 24 months of the STMP start date with no periods excluded from the estimation sample.

2. Examine offending within 12 months of the STMP start date excluding a six-month interval centred on the STMP start date (i.e., excluding three months on either side) from the estimation sample.

3. Examine offending within 24 months of the STMP start date excluding a six-month interval centred on the STMP start date (i.e., excluding three months on either side) from the estimation sample.

4. Examine offending within 24 months of the STMP start date excluding a 12-month interval centred on the STMP start date (i.e., excluding six months on either side) from the estimation sample.

5. Use the matched control group (outlined in the Appendix of this document and in the Appendix of Yeong (2020)) in a Difference-in-Differences (DID) setup.

Estimates generated from approaches (1) - (5) are, respectively, reported in columns 1 - 5 of Table A. Approach (1) is provided for completeness. The intuition behind approaches (2) - (4) is to remove the increase and decrease in crime occurring immediately before and after the STMP start date, respectively. Said differently, these robustness checks remove the hump from the estimation sample and then re-estimate the model employed by Yeong (2020). It should be noted that these robustness checks are in no way definitive; they simply allow us to determine whether the estimates reported by Yeong (2020) are entirely driven by the hump over the periods of time excluded from the estimation sample in each robustness check. In approach (5) we employ a DID approach that compares individuals subject to STMP, with the matched control group (from Yeong (2020)), before and after STMP. This approach differs from Yeong's primary approach in that we use individuals never subject to STMP as a control for individuals subject to STMP. This approach also differs from the matched comparison reported in Table A2 of Yeong (2020). In Table A2, Yeong (2020) compared the likelihood of offending between groups after placement on STMP. The DID approach on the other hand, takes the difference in offending risk prior to STMP into account when estimating the association between STMP and crime. The validity of the DID approach is contingent upon the assumption that the

[1] The specific set of factors considered for STMP-II is outlined in footnote 4 of Yeong (2020). These factors include: whether the individual was involved in crime at a young age; prior offending; whether the use of violence and/or a weapon was involved in such offences; prior sentences of imprisonment; prior community orders; whether the individual in question has any known criminal associations, addition and mental health issues. While these factors generate a risk score that is used to guide the discussion around whether an individual is placed on STMP, ultimately whether an individual is placed on STMP is determined by police discretion.

matched control group is able to provide a valid counterfactual outcome for the treatment group in the absence of STMP. In the Appendix of this document we find evidence to indicate that this assumption does not hold. As such, this robustness check offers little more than the primary approach employed by Yeong (2020). Nonetheless, we report the estimates from this robustness check for completeness and consistency with prior work (Machin and Marie, 2011).

Panels A, B and C of Table A examine the relationship between STMP-II and the probability of at least one selected violent or property crime.[2] Panel D examines the relationship between DV-STMP and the probability of at least one DV offence.

Panel A examines all individuals subject to STMP-II. The estimate in column 1 is negative and statistically significant, although a little smaller than its counterpart in Table 2 of Yeong (2020). This indicates that as we increase the time span of the estimation sample, the association between STMP-II and crime remains but weakens. There are at least two explanations for this. The first is that the association is being driven by the hump, and thus as we increase the estimation sample, the contribution of the hump dissipates. The other is that STMP only has a short-term association with crime. This could be because the level of police supervision diminishes as an individual's proclivity for crime falls through deterrence or incapacitation (assuming that STMP does reduce crime) or individual's perceive a lower level of apprehension risk as their experience with the program grows. These explanations are not mutually exclusive.

From columns 2 - 4 of Panel A, we can see that the association between STMP-II and at least one subsequent violent or property crime is approximately a 2.5 percentage point reduction. This is true even when we exclude an entire year of data around the STMP start date. In terms of absolute magnitude[3], the estimates in Table A are around one-third the size of the estimates reported by Yeong (2020). In terms of relative magnitude[4], the associations in Panel A are between one-third to one-half the relative reductions reported by Yeong (2020). From column 5, we can see that the point estimate is largely consistent with its counterpart in Tables 2 and 3 of Yeong (2020).

Panel B examines juveniles subject to STMP-II. Again, the estimate from column 1 is a little smaller than its counterpart in Yeong (2020) but remains negative and statistically significant. From columns 2 - 4 of Panel B we can see that the association between STMP-II and at least one selected crime is around a 5.3 percentage point reduction. In terms of absolute magnitude, this is again about one-third the size of the estimates reported by Yeong (2020). In terms of relative magnitude, the associations in Panel B are between one-third to one-half of the relative reductions reported by Yeong (2020). The estimate in column 5 is largely consistent with its counterpart in Yeong (2020).

Panel C examines Aboriginal people subject to STMP-II. The estimate in column 1 is about one-half the size of its counterpart in Table 2 of Yeong (2020) and statistically significant at the five per cent level. The estimates in columns 2 - 4 are all statistically insignificant. This indicates that there is likely a sizable degree of mean reversion occurring in relation to the estimates reported by Yeong (2020) for Aboriginal people. From column 5 we can see that, if anything, the point estimate is actually larger than its counterpart in Yeong (2020). This is likely due to the fact that one of the key variables Yeong (2020) failed to adequately match on was Aboriginality.

Panel D of Table 1 examines individuals subject to DV-STMP. The estimate reported in column 1 is similar to its counterpart in Table 3 of Yeong (2020). From columns 2 - 5 of Panel D we can see that the association between DV-STMP and DV crime is around a 21 percentage point reduction. In terms of absolute magnitude, this is approximately two-thirds of the size of the estimate reported by Yeong (2020). In terms of relative magnitude, the associations in Panel D are broadly consistent with those reported by Yeong (2020).

To summarise, the robustness checks reported in Table A indicate that the estimates reported by Yeong (2020) were likely subject to some degree of mean reversion. That said, with the exception of the estimates for Aboriginal people, all of the estimates reported in Table A are consistent with the primary finding of Yeong (2020): that STMP is associated with a reduction in crime. The estimates with regard to Aboriginal people, however, warrant further investigation. While the estimates are all negative once periods immediately

[2]As outlined in Yeong (2020), we focus on a selected subset of violent and property crimes to avoid reporting/detection bias (i.e., the idea that once placed on STMP, an individual is more likely to get caught, irrespective of whether their actual offending rate changes). These crimes include: homicide, assault occasioning grievous bodily harm, robbery, theft, motor vehicle theft and break and enter

[3]That is, comparing the size of the coefficients in Table A with those in Table 2 of Yeong (2020)

[4]That is, expressed as a fraction of the unconditional pre-STMP probability of offending.

preceding and following STMP are removed from the sample, they are not statistically (or practically) significant. Interpreting this result is difficult. On one hand, as Machin and Marie (2011) point out, discarding six or 12 months of data is a very stringent test. On the other hand, however, this result does beg the question of whether the estimates reported by Yeong (2020) were the result of regression to the mean.

Table A: Robustness checks

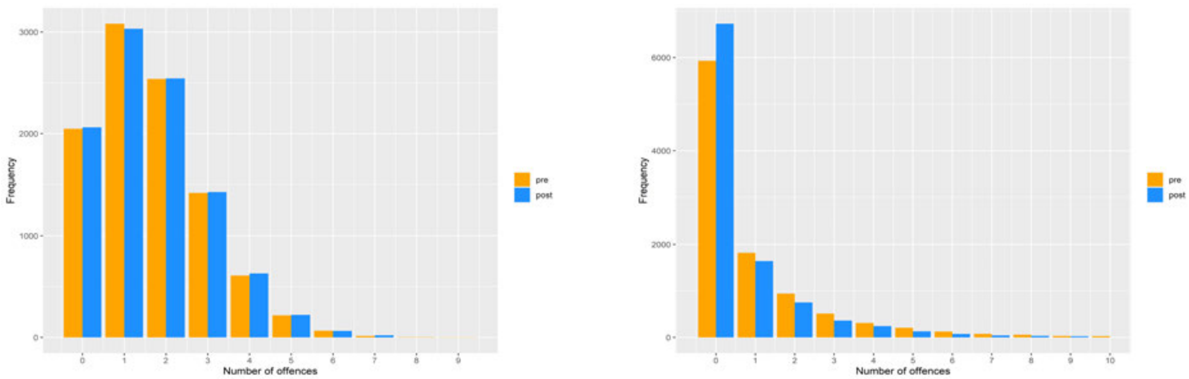
	(1) 24 month follow up	(2) Six month interval with 12 month follow up	(3) Six month interval with 24 month follow up	(4) 12 month interval with 24 month follow up	(5) Matching + DD
Panel A.					
Everyone on STMP-II	-0.052***	-0.024***	-0.029***	-0.023***	-0.077***
	(0.007)	(0.006)	(0.007)	(0.007)	(0.008)
Pre-policy mean	0.522	0.272	0.417	0.356	0.414
Observations	20,120	20,120	20,120	20,120	37,612
Adjusted R-squared	0.126	0.069	0.102	0.089	0.183
Panel B.					
Juveniles on STMP-II	-0.120***	-0.061***	-0.057**	-0.040*	-0.150***
	(0.017)	(0.015)	(0.018)	(0.018)	(0.019)
Pre-policy mean	0.744	0.454	0.591	0.500	0.667
Observations	5,460	5,460	5,460	5,460	7,788
Adjusted R-squared	0.080	0.039	0.068	0.069	0.223
Panel C.					
Aboriginal Australians on STMP-II	-0.026*	-0.015	-0.009	-0.009	-0.073***
	(0.010)	(0.009)	(0.010)	(0.010)	(0.012)
Pre-policy mean	0.585	0.318	0.481	0.417	0.466
Observations	9,192	9,192	9,192	9,192	18,140
Adjusted R-squared	0.118	0.069	0.097	0.086	0.190
Panel D.					
Everyone on DV-STMP	-0.319***	-0.201***	-0.250***	-0.177***	-0.244***
	(0.022)	(0.020)	(0.022)	(0.024)	(0.027)
Pre-policy mean	0.797	0.480	0.603	0.447	0.727
Observations	2,050	2,050	2,050	2,050	3,858
Adjusted R-squared	0.147	0.069	0.101	0.072	0.268
Follow up time	24 months	12 months	24 months	24 months	12 months
Period excluded	None	3 months on either side	3 months on either side	6 months on either side	None

Note. PAC = Police Area Command, all estimates include the control variables described in Yeong (2020) and PAC-by-year fixed effects, robust standard errors in parentheses, *** $p<0.001$, ** $p<0.01$, * $p<0.05$.

Before moving on, it should be noted that applying the robustness checks in columns 1 - 4 of Table A to the simulation reported by Popovic (2020) still results in a negative relationship between STMP and crime. While we agree that this simulation is useful in illustrating the issue associated with Ashenfelter's Dip, it is important to bear in mind that this simulation does not reflect the actual selection process for STMP, nor the relationship between STMP and crime. In particular, the simulations described by Popovic (2020) and Macdonald (2020) assume that placement on STMP is purely a function of prior offending. In practice, we know that this is not true. Conceptually, as reported by Yeong (2020), placement on STMP is a function of prior offending (for a specific set of offences), police discretion, prior prison sentences, community orders, whether the individual was involved in crime as a juvenile, has any known criminal associations, mental health and/or addiction issues. Empirically, we know that controlling for an observable subset of these factors makes a sizable difference to the estimates.

The fact that the simulations don't take these factors into account makes a comparison between the actual data and the simulations difficult. For example, in Popovic's simulation, around 95 per cent of individuals have at least one offence in the 12 months prior to placement on STMP. However, we know from Table 2 of Yeong (2020) that this number is closer to 40 per cent. This difference between these pre-STMP offending probabilities is driven by Popovic's data generating process. In Popovic's simulation, all individuals commit 20 offences, on average each individual has *lambda* days between any two offences, and one of these (pre-STMP) offences triggers placement on STMP, which occurs in *wait* days after the triggering offence. Importantly, both *lambda* and *wait* are conditional parameters. The parameter *lambda* requires an individual to have at least two offences and *wait* requires at least one. Given that 60 per cent of the sample reported by Yeong (2020) have zero selected violent or property offences in the 12 months before STMP, this simulation is difficult to reasonably reconcile with the actual data.

Furthermore, there are key differences between the simulated and actual data in offending frequency after placement on STMP. In Popovic's simulation, for example, the distribution of the offence counts after the program start date is very similar to that observed before the program commences (as shown in Panel A of Figure A below). This is because STMP has no effect in her simulation. In the actual data, however, the pre- and post-STMP offence count distributions are quite different. In total, there were 10,668 selected violent and property offences in the 12 months before STMP-II, compared to 8,030 offences after STMP-II. This represents an average of 0.261 fewer offences post-STMP in the actual data. The same number for the simulated data is only 0.007, which is consistent with the simulation setup in which there is no policy effect.



(a) Simulated data

(b) Actual data

Figure A: Panel A reports offence counts based on Popovic's simulated data. Panel B reports offence counts based on the actual data from Yeong (2020).

Matching

Given the purported limitations of the approach employed by Yeong (2020), Watson (2020) suggests that a matching strategy would be more appropriate. A matching strategy was employed by Yeong (2020) and the results were reported in Tables A1 and A2 of the Appendix. Intuitively, this approach involved searching through BOCSAR's Reoffending Database[5] for a set of individuals (never placed on STMP) that were similar to individuals placed on STMP, proximate to the time that they (i.e., the treated individual within each matched pair) were placed on the program. In order to identify such individuals, Yeong (2020) attempted to match on a set of (observable) individual level characteristics relevant to selection for STMP.

Yeong (2020) was, however, unable to find a set of similar individuals to use as a control group. For example, in terms of observable characteristics, individuals in the control group were older, more likely to be Aboriginal and have more prior offences than people actually subject to STMP. It is, however, the unobservable characteristics that are the real cause for concern. As outlined by Yeong (2020), there are a variety of factors explicitly considered when determining whether to place an individual on STMP that also influence offending (i.e., police discretion, known criminal associations, mental health and addiction issues). Failure to account for such factors helps to explain the results Yeong (2020) found in Table A2 of the Appendix: that offenders placed on STMP are more likely to offend than their matched counterparts.[6]

To better understand this result, consider the two endogeneity problems outlined in Yeong (2020): reporting/detection bias; and selection bias. The reporting/detection bias issue refers to the idea that individuals placed on STMP are more likely to get caught once they become subject to increased police supervision. To

[5]Which contains information for all individuals with a finalised court appearance between January 1994 and September 2019.

[6]The results reported in Table A2 of Yeong (2020) differ from the results reported in column 5 of Table A in this document. This is because the DID estimates account for pre-existing differences in the probability of offending prior to placement on STMP, while the estimates reported in Table A2 of Yeong (2020) do not.

address this issue, consistent with prior research outlined in the literature review, Yeong (2020) focused on a specific set of violent and property offences. The selection bias problem refers to the issue that individuals placed on STMP are at higher risk of offending than other individuals known to police. Absent the existence of: a) a group of offenders that were not placed on STMP 'by chance'; and b) the capacity to observe all factors that influence selection for STMP and offending, matching cannot address the selection bias issue. As such, left unaddressed, we would expect the control group to exhibit a lower rate of offending than those subject to STMP.[7] This is illustrated in Figure B in the Appendix of this document, where we can see that individuals placed on STMP offend at much higher rates than their matched counterparts, both *before and after* placement on STMP.

Nonetheless, Watson (2020) maintains that a different matching strategy (e.g., genetic matching) may have achieved superior balance across groups and should have been more vehemently pursued. On page 10 of his review, Watson (2020) states that:

> *'as long as the researcher achieves good balance on the covariates between treatment and control groups, then regression modeling may proceed with reasonable confidence.'*

Once again, this is not correct for the reasons described above. Another way to make this point is to say that the assumption underpinning all matching strategies (i.e., conditional independence between the treatment variable and the error term (Angrist and Pischke, 2008)), is violated in the context of the Yeong (2020) study.

Reporting of the descriptive statistics.

Another criticism from Watson (2020) is that Yeong (2020) reports means instead of medians in the descriptive statistics section of the paper. Here it is important to bear in mind that the purpose of descriptive statistics is to help the reader: a) understand the data; and b) interpret the models. Given that Yeong (2020) estimates models primarily using ordinary least squares, reporting means is a more consistent approach to inform the reader of what the model is estimating.[8] As to whether means are useful in understanding the data, Table B of this document reports the mean and median associated with each of the variables reported in Table 1 of Yeong (2020). From Table B, we can see that there is not a substantial difference between these two statistics.

[7] In fact, once we attempt to account for this selection bias through the use of a DID model, the estimates become negative and largely consistent with the main results reported by Yeong (2020).

[8] Yeong (2020) also reports estimates using competing nonlinear models (in Table A4 of the Appendix).

Table B: Means and medians

	One year before STMP		First day on STMP	
	Mean	Median	Mean	Median
	(1)	(2)	(3)	(4)
Panel A. STMP-II				
Age	24.944	23	25.961	24
Age at first CJS contact	17.941	16	17.941	16
Male	0.915	1	0.915	1
Aboriginal	0.458	0	0.458	0
Prior court appearances	8.210	6	9.703	8
Prior violent offences	3.617	2	4.500	3
Prior weapon offences	0.396	0	0.544	0
Prior drug offences	1.309	0	1.661	1
Prior community orders	4.704	4	4.704	4
Prior YJCs and cautions	0.526	0	0.692	0
Prior prison sentences	1.741	0	1.995	0
Panel B. DV-STMP				
Age	34.031	33	35.048	35
Age at first CJS contact	22.003	19	22.003	19
Male	0.914	1	0.914	1
Aboriginal	0.371	0	0.371	0
Prior court appearances	9.529	7	10.924	9
Prior violent offences	6.675	5	8.894	7
Prior weapon offences	0.408	0	0.505	0
Prior drug offences	1.278	0	1.523	1
Prior DV offences	4.371	2	8.129	6
Prior community orders	4.340	4	4.340	4
Prior YJCs and cautions	0.403	0	0.420	0
Prior prison sentences	1.589	0	1.871	1

Causal claims

Throughout the report, Yeong (2020) attempted to make clear that the estimates do not have a causal interpretation. For example, an extract from the method section (paragraph 5, page 8) reads:

> *'In order for β_1 to have a causal interpretation, an individual's risk of offending must be conditionally unrelated to this timing. Given that the timing of when an individual becomes subject to STMP is a direct function of their offending behaviour, there is no reason we should expect this condition to hold.'*

Yeong (2020) then goes on to say (in paragraph 1, page 9):

> *'my estimates do not have a causal interpretation. Instead, they must be interpreted as the association between STMP and offending.'*

And then again in the discussion (in paragraph 2, page 18) Yeong (2020)) says:

> *'The present study is not, however, without its caveats. The most important of which is that the estimates do not have a causal interpretation. That is, because I am simply comparing the behaviour of individuals before and after placement on STMP, I have no way of establishing what would have happened in the absence of STMP.'*

That said, we acknowledge that in some sections of the report and in the one page summary there was a causal 'tone' to the language used. Revisions have since been made to both these documents to minimise the risk of the results being misrepresented.

Technical issues

Another criticism raised by Watson (2020) was the relatively low adjusted R-squared for the models utilised by Yeong (2020). Models with a relatively low R-squared are common in applied micro-econometrics, where the focus is typically on causal inference, not prediction (Angrist and Pischke, 2008). In applied work where casual inference is the focus, the key insight the R-squared offers is around how one should interpret the stability of a coefficient in the face of control variables (Oster, 2019). In addition to the adjusted R-squared, however, Yeong (2020) also reported the values for the Area Under the receiver operator characteristic Curve (AUC) when checking the robustness of his results against competing nonlinear models (in Table A6). The AUC for all of the Probit models reported by Yeong (2020) are within the acceptable range, which Mandrekar (2010) characterises as 0.7 to 0.8.

Another issue raised by Watson (2020) is that Yeong (2020) should have used a multilevel model (or Random Effects estimator) to address questions around possible treatment effect heterogeneity (i.e., the idea that STMP has a different impact for different groups of individuals). There are two problems with this suggestion. The first is that coefficients from a consistent Random Effects model will necessary produce a similar coefficient of interest to the Fixed Effects model used by Yeong (2020). Said differently, even if the restrictive set of assumptions underpinning the Random Effects approach were satisfied, the magnitude and direction of the primary estimates reported by Yeong (2020) would remain unchanged. The second issue relates to whether or not these assumptions are valid. The Random Effects estimator requires that the unobserved heterogeneity (i.e., the PAC effect) be unrelated to the control variables included in the regression (e.g., Aboriginality, prior community orders, offending and prison etc) (Wooldridge, 2010). Intuitively, it is easy to see how (geographically defined) police jurisdictions may be correlated with the likelihood that an individual identifies as Aboriginal (as some communities have larger Indigenous populations) and prior offending (as some communities have higher rates of crime than others). In order to safeguard against this issue, Yeong (2020) takes the more conservative approach and employs the Fixed Effects estimator.

Conclusion

Evaluating the causal impact of the STMP program on crime is challenging. The program necessarily targets high-risk offenders who differ from other groups on observable (and likely unobservable) characteristics. Given this, and the way in which the program was rolled out, there is no natural control group for program participants. This makes it very difficult to identify a valid counterfactual. As such, we acknowledge that the identification strategy used by Yeong (2020) does not allow for a causal estimate because selection into the treatment group is (in part) conditional on the outcome. However, the evidence presented above, on balance, supports the conclusion from Yeong (2020): that STMP has a negative association with crime. The exception is for Aboriginal Australians for whom the evidence of a negative association is much weaker. As argued by Yeong (2020), this result, combined with the significant increased risk of imprisonment associated with the program, indicates that STMP may need to be reviewed for this particularly vulnerable group.

References

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press, illustrated edition edition.

Ashenfelter, O. (1978). Estimating the Effect of Training Programs on Earnings. *The Review of Economics and Statistics*, 60(1):47–57.

Ashenfelter, O. and Card, D. (1985). Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. *The Review of Economics and Statistics*, 67(4):648–660.

King, G. and Nielsen, R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, 27(4).

Macdonald, J. (2020). Demonstrating an illusory reduction in crime via sampling.

Machin, S. and Marie, O. (2011). Crime and Police Resources: The Street Crime Initiative. *Journal of the European Economic Association*, 9(4):678–701.

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*, 5(9):1315–1316.

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Popovic, G. (2020). Simulation experiment for "An evaluation of the Suspect Target Management Plan, October 2020" Study.

Watson, I. (2020). A critical review of the BOSCAR report: An evaluation of the Suspect Target Management Plan.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data, second edition*. MIT Press Academic, Cambridge, Mass, 2 edition edition.

Yeong, S. (2020). An evaluation of the Suspect Target Management Plan (STMP). *Crime and Justice Bulletin, 233*, NSW Bureau of Crime Statistics and Research.

Appendix

This Appendix has two parts. The first part provides an overview of the matching procedure employed by Yeong (2020). The second part outlines the Difference-in-Differences (DID) approach used in this document.

The matching procedure used by Yeong (2020)

In order to identify a suitable control group for individuals subject to STMP, Yeong (2020) used a quarterly individual level panel (i.e., one row per individual-quarter-year combination). This dataset contains information for each of the variables outlined in Table 1 of Yeong (2020) and Table B of this document, for any individual with a finalised court appearance between 1 January 1994 to 31 September 2019. Each variable in the dataset is indexed to the first day of a given quarter-year (e.g., age and number of prior offences for each individual as of 1 January 2010 for Q1-2010, 1 April 2010 for Q2-2010 and so on).

Using this dataset, Yeong (2020) then employed the following matching procedure:

1. Limit the estimation sample to a given quarter-year (e.g., 2010-Q1).
2. Retain individuals that either: began STMP within the given quarter-year; or were never subject to STMP.[9]
3. Use Coarsened Exact Matching (CEM) to further limit the sample to the subset of treatment-control observations within the area of common support.
4. Use Propensity Score Matching (PSM) to obtain the best possible 1:1 match between each treatment and control unit.
5. Recalculate all variables to the day that the treated unit began STMP (within each matched pair).[10]
6. Repeat steps 1 - 5 for each quarter-year between 2005-Q2 and 2018-Q3.

[9] Potential control units that were in custody on the first day of the quarter are excluded from donor pool.

[10] That is, because all of the variables are indexed to the first day of a given quarter-year, Yeong (2020) needed to re-index these variables to the day that the treated unit began STMP. For example, within a given pair, if the treated unit began STMP on 1 February 2010, in step 5, Yeong (2020) recalculated both the treated and control unit's variables (e.g., reoffend within 12 months, number of prior offences) as of 1 February 2010 (instead of the first day of 2010-Q1, which is 1 January 2020).

This matching algorithm has several advantages over using CEM or PSM alone. As outlined by King and Nielsen (2019), the use of CEM before PSM safeguards against extrapolations made in PSM that can lead to model dependant inferences. King and Nielsen (2019) also argue that PSM, used in conjunction with CEM, is potentially better than CEM alone in situations with a large degree of imbalance between groups. Said differently, PSM (after CEM) works well in circumstances where causal inferences are least likely because the treatment and control groups are so different. Another advantage to pre-processing the data using CEM is that CEM is computationally efficient, which is of practical importance given that Yeong (2020) had, for each quarter-year, over 1.3 million potential control units.

The Difference-in-Differences model used in this document

The Difference-in-Differences approach used in this document is summarised in Equation A1 below.

$$y_{it} = \beta_0 + \beta_1 * (treatment_i * post_t) + treatment_i + post_t + \gamma \mathbf{X'_{it}} + \epsilon_{it} \tag{A}$$

Where y_{it}, $post_t$, $\mathbf{X'_{it}}$ and ϵ_{it} all have the same definition as in Equation 1 of Yeong (2020)[11]; $treatment_i$ is a binary variable equal to one for individuals subject to STMP, zero for individuals in the matched control group (from Yeong (2020)); and all other terms are coefficients to be estimated.

In Equation A, the coefficient of interest, β_1, represents the association between the probability of at least one offence and STMP. The difference between Equations 1 and A is that, in Equation A, we are using non-STMP participants as a control group for individuals subject to STMP. The idea is that if the difference in the (un)observable characteristics between treatment and control groups remains constant over time, then Equation A1 should difference out (remove) the selection bias.

If Equation A is able to address the selection bias issue, we would expect the treatment and control groups to share parallel trends in the evolution of offending prior to STMP. This proposition is examined in Figure B, which plots the daily probability of at least one selected violent or property crime within 12 months of STMP-II.

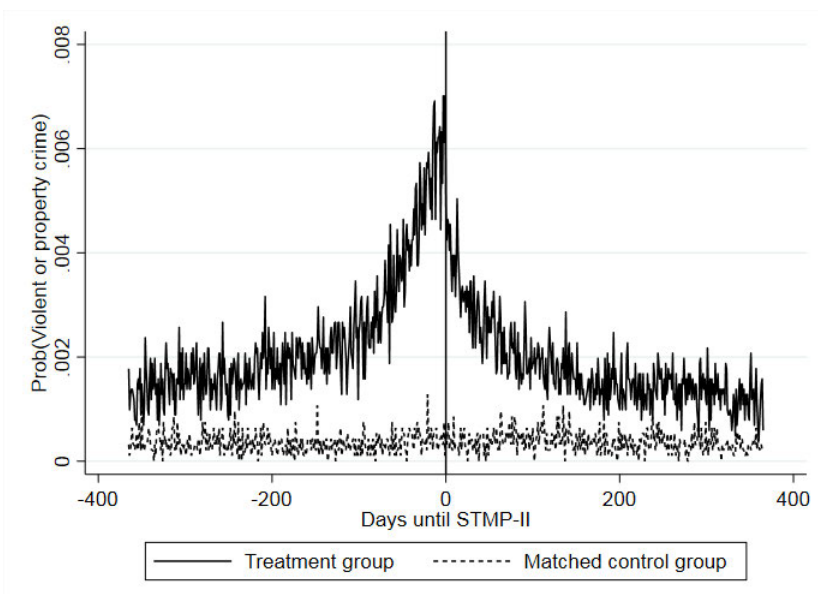


Figure B: Treatment vs matched control group

There are a few observations of note with respect to Figure B. First, individuals subject to STMP are of a higher risk of offending than their matched counterparts. Given that this is true both before and after STMP, this explains why Yeong (2020) found a positive relationship between STMP and offending in the Appendix. As Yeong (2020) points out on page 24:

[11] Although we are no longer indexing by PAC.

> *‘One explanation for this finding is that there is some form of unobserved heterogeneity that matching cannot address. For example, known criminal associations, addiction issues and police intelligence are important unobserved factors likely to influence program participation.’*

The second observation of note with respect to Figure B relates to the capacity for the DID approach to address the selection bias problem. Recall that the DID model’s capacity to address such an issue depends on whether the difference in offending between groups remains constant over the pre-STMP period (i.e.., whether the two groups share parallel trends prior to STMP). From Figure B we can see that this is not the case. The risk of offending in the treatment group is clearly increasing in the lead up to STMP. The risk of offending for the control group appears to be independent of STMP. This is despite Yeong’s attempt to find a suitable match for individuals subject to STMP proximate to the time that they were placed on STMP. Said differently, if Yeong’s matching procedure did enable him to find a suitable match for individuals subject to STMP (proximate to the date that they were placed on the program), we would expect to see an upward pre-policy trend in the control group’s risk of offending. However, the fact that there is no trend in offending for the control group indicates that this DID approach offers little more than the much simpler pre vs. post comparison employed by Yeong (2020).