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

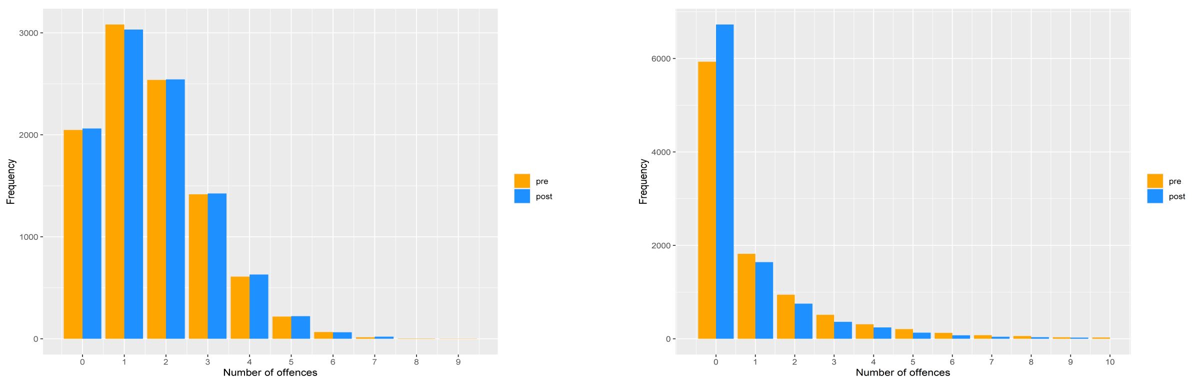
	(1) 24 month follow up	(2) Six month interval with 12 month follow up	(3) Six month interval with 24 month follow up	(4) 12 month interval with 24 month follow up	(5) Matching + DD
Panel A.					
Everyone on STMP-II	-0.052***	-0.024***	-0.029***	-0.023***	-0.077***
	(0.007)	(0.006)	(0.007)	(0.007)	(0.008)
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.

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

Where y_{it}, $post_t$, $\mathbf{X}'_{\mathbf{it}}$ 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.

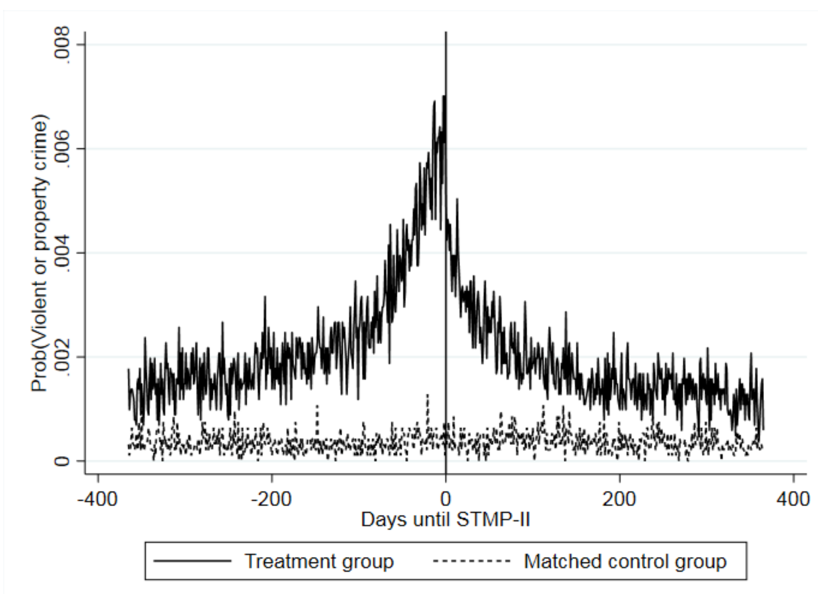


Figure B: Treatment vs matched control group

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).