

# Checkpoint: An innovative Programme to navigate people away from the cycle of reoffending – A randomised control trial evaluation

The Police Journal:  
Theory, Practice and Principles  
1–28

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0032258X211018774

journals.sagepub.com/home/pjx



**Kevin Weir, Stephanie Kilili , John Cooper, Andrew Crowe and Gillian Routledge**  
Durham Constabulary, Aykley Heads Durham, Durham, UK

## Abstract

This study contributes to the evidence base of police deferred prosecution schemes aimed at reducing reoffending. Durham Constabulary, UK, introduced Checkpoint, an adult deferred prosecution scheme which targets offenders entering the Criminal Justice System by providing an alternative to a criminal prosecution. Applying theories of deterrence and desistance, this paper describes the randomised control trial findings of 521 offenders randomised between 1 August 2016 and 31 March 2018. The results indicate that the Checkpoint treatment cohort achieved a lower reoffending rate in comparison to the control cohort, on the basis of prevalence (10.3% reduction) and risk of reoffending (30% reduction).

## Keywords

Reoffending, out of court disposals, evidence-based policing, offender desistance

## Introduction

Out-of-Court Disposals (O OCD) have become a key policy and political area in the UK, with the removal in April 2013 of the previous requirement on the Crown Prosecution Service to decide on the suitability of cases for conditional cautions, devolving the decision to the custody officer in the police station (Neyroud and Slothower, 2013).

---

## Corresponding author:

Stephanie Kilili, Durham Constabulary, Durham Police HQ, Aykley Heads, Durham DH1 5TT, UK.

Email: stephanie.kilili@durham-pcc.gov.uk

This created not only an increase in time spent on a disposal decision by the police but also a requirement on the police to set and manage conditions for offenders. Some were of poor quality or not applied correctly, and invariably, no attempt was made to understand what drove offending behaviour (CJJI, 2011).

In response to the 2011 Criminal Justice Joint Inspection report (CJJI), the Ministry of Justice (MoJ) initiated a consultation with the public and police practitioners with a view to reforming the adult disposal framework and legislative and procedural changes to OOCs. In October 2017 DCC Glen, the National Police Chiefs' Council lead for charging and OOCs, submitted her national strategy for consideration to the Chief Constables' Council; this strategy sought to move policing voluntarily from the existing multiple choice framework towards a two-tier framework, based on a community resolution (CR) and a conditional caution (CC). While the reform does indeed represent a simplification to a more straightforward, escalatory process, the onus of the evaluation was on the implementation of the new framework rather than on its effectiveness. The sparsity of quantitative studies in terms of comparing diversionary and conventional approaches, using offender cohorts that share similar characteristics, remains a clear and present issue for OOC policy.

In the UK, only a limited number of studies have taken place to assess the effectiveness of cautions and diversion but for those that have, they suggest that diversion may well be more effective than formal processing (Neyroud and Slothower, 2013). One of the first attempts in the UK to test a diversionary approach against a normal prosecution procedure was Operation Turning Point (OTP) run by West Midlands Police (Neyroud and Slothower, 2013). The overall aim of OTP was to 'test whether low harm offenders, who might otherwise have been prosecuted, could be dealt with in a combination of a deferred prosecution and tailored conditions to encourage desistance' (Neyroud and Slothower, 2013: 22). Results showed increased victim satisfaction, significant cost savings but less conclusive findings in relation to harm and reoffending frequency (Neyroud et al., 2015). In 2015, Durham Constabulary adopted a deferred prosecution model, namely Checkpoint, in order to assess its impact on reducing reoffending and improving offenders' life chances (Weir et al., 2019). The main differences between the two can be found in Table 1.

Checkpoint has been designed to take a public health approach to policing based on the concept of Offender Desistance Policing (ODP) (Sherman, 2011; Sherman and Neyroud, 2012). The objective of the intervention is to reduce reoffending by giving offenders the opportunity to address the underlying causes of their offending behaviour and improve life chances. Checkpoint is an offender management programme that seeks to address the underlying reasons why someone has offended and places a greater degree of accountability on the offender to face up to the consequences of their offending behaviour, by working with an offender manager or Navigator on a one-to-one basis. Offenders receive the support they need to address the reasons they offend, acknowledging the impact of their actions, to make reparations for the harm they have caused and helping them to move on with their lives.

One of the key criticisms of research into OOCs has been that, despite the fact that such disposals have been in use for decades, little in the way of testing OOCs against prosecution processes has taken place using appropriate research methods (Neyroud,

**Table 1.** Main differences between turning point and checkpoint.

	<b>Turning Point</b>	<b>Checkpoint</b>
<b>Delivery of intervention</b>	Police Officers	Specialist, independent Navigators
<b>Eligibility Criteria</b>	First-time Offenders	Offenders with less than three previous offences
<b>Included Offences</b>		Additional offences included
<b>Location</b>	One pilot site	Entire police force area
<b>IT System</b>	No new IT system	New, bespoke in-house IT system called COMET to store and manage offender needs assessments and contract conditions.

2018). Therefore, to address this deficit in policing research, the evaluation of the Checkpoint treatment programme was based upon an experimental research design, in the form of a randomised controlled trial (RCT). The results of the research need to be robust enough to secure the confidence of not only an internal audience, that being Durham Constabulary and its partners contributing time and funding to the project, but to an external audience of UK police forces and, perhaps most importantly, government policy makers.

The theoretical premise for Checkpoint derives from ‘Offender Desistance Policing (ODP)’ (Sherman, 2011) which looks at how the police can support the rehabilitation of offenders by diversion better than the processes of prosecution, sentencing and punishment. ODP makes an appeal to deterrence theory and desistance processes to achieve this aim.

Looking first at the appeal made to deterrence theory by ODP, there are a number of theories and studies which provide a foundation on why people offend and what can influence them to stop offending (Farrall and Calverley, 2005). Thomas Hobbes (1651) and Beccaria (1767) for example suggested that if people have their own rational self-interest at heart then this would lead to crime, and invariably, conflict and exclusion from society. Hobbes advocated that even when the state intervened, individuals would still offend, and therefore swift and certain punishment rather than severity were the best ways to deal with criminality (1994). Hobbes also stated that the punishment for the crime must out-weigh its perceived benefits and rewards (1994). Based on this, the deterrent effect in Checkpoint must be swift; there must be an imminent threat of prosecution for those who fail to comply or reoffend, and the punishment for non-compliance, i.e. receiving a criminal record must out-weigh the benefit of the crime itself.

Research suggests a developing consistency for the deterrent effect regarding the certainty of punishment being greater than the severity of punishment (Klepper and Nagin, 1989; Paternoster, 1987; Williams and Hawkins, 1986). Parker and Grasmick (1979) suggest that the threat of legal sanctions is only successful for a proportion of the population as some individuals are beyond correction or reform, with any deterrent effect likely to vary depending on social position. On the basis that it may be the imminent threat of punishment, not the severity that deters people, Sherman and Neyroud illustrate this approach using the metaphor of the ‘Sword of Damocles’ (2012). In the context of

Checkpoint, the threat of imminent prosecution hanging over the offender, like the sword hanging over Damocles, rather than the prosecution itself, is expected to have a focussed deterrent effect. Kennedy et al. (2017), emphasise the importance of certainty over severity but also stress the importance of the police delivering a message that they can later follow through on. Conceptualising this, if people who have offended successfully comply with their Checkpoint 'contract to engage', then the police will follow through with their commitment to drop the relevant sanction and not label the individual as an offender.

Moving on to the contribution of desistance processes to ODP, desistance is not a new concept, yet how and why offenders change has been under-researched (McNeill, 2002). Desistance is generally divided into three areas:

'Maturational reform' theories are based on the maturity and criminal behaviour of an individual. Some of the most cited academic research is on the 'turning points' by Sampson and Laub (1997) who developed the idea that offenders will offend less as they mature in age.

'Social Bond' theories are concerned with the whole life course and suggest that where an individual who has attachment to family, friends and marriage, suitably paid employment, then having these structured bonds can change the criminality of an individual. Farrington (2003) also proposes that labelling someone as an offender, poses a risk of developing social norms.

'Narrative theories' have developed from qualitative research, which is personal to the individual, such as the self-identify, concerns for others and a greater consideration for the future. Giordano et al. (2002) examined the social learning perspective across the life course, focusing on what changed and capturing how an individual relates to their 'roles and behaviours'. Desistance, in this sense, is about an individual 'fearing one's self' and the situation they find themselves in, with a view to shaping a new positive self (Oyserman and Markus, 1990).

The importance of the desistance process in ODP is exemplified by the work of the offender manager, or Navigator in the context of Checkpoint, and their work with the offender. A study which involved interviews with probation officers and their offenders suggests that when probation officers had acted as a role model giving the offenders a 'sense of obligation' by reinforcing pro-social behaviour, the reoffending rate was less (Rex, 1999). Trotter also proposed that it may be the social worker skills and problem-solving that achieved these outcomes (1996, 2000).

The criminogenic needs of offenders are covered in many academic papers but it wasn't until the 1990s that the criminogenic needs of an offender were examined to see what worked in reducing recidivism and supporting desistance (Andrews et al., 1990a, 1990b; Gendreau et al., 1996). Reducing recidivism rates has been shown to be more effective using community-based interventions rather than punishment through imprisonment (Gendreau and Ross, 1981). The factors aligned with the most successful programmes include:

- i. The period was usually for a few months in duration, focusing on social learning and the conceptualisation of criminal behaviour;
- ii. Uses cognitive skill and targets the criminogenic needs of offenders;

- iii. The interventions are delivered in a way to enable the offender to learn new pro-social skills;
- iv. Programmes are firm, fair and have a positive reinforcement;
- v. The offenders are interacted in an interpersonal and sensitive way;
- vi. Programmes reach out to the offender's social network and target their disruptive activities with pro-social ones.

While research on what works to stop offending has focussed on post-court interventions, very little research exists on pre-court interventions and therefore this study is important in increasing the evidence base for use of police deferred prosecution schemes. A review by Blakeborough and Pierpoint (2007) of six UK police forces identified that conditional cautioning schemes helped to address the causes of crime. Checkpoint adopts a similar approach, with the main difference being that if an individual successfully completes, then they are not labelled as an offender at the end of the conditions and do not receive a conviction at court. A meta-analysis by Harvey et al. (2007) reviewed studies across North America and Australia involving deferred prosecution and diversion and noted that a vast majority of these studies evidence a statistical difference in reduced reoffending.

In summary, in contextualising the criminological theory, it suggests that the threat or fear of punishment is far more effective than any criminal sanction. As such, Checkpoint should use the threat of punishment, or the 'Sword of Damocles' (Sherman and Neyroud, 2012), to gain compliance within a short and tailored programme. Checkpoint should use specialist 'Navigators' to act as role models and gain an understanding of the offender to address their criminogenic needs. Weir et al. (2019), describe in greater detail how the work of the Navigators with offenders over the duration of the Checkpoint contract chimes with the life-course criminology. Employing problem-solving approaches, removing the label of being a criminal and providing an opportunity to improve their life chances is hypothesised to more likely assist the offender to successfully complete the intervention. If the Checkpoint scheme is designed with these in mind, it is expected to be more likely to lead to reductions in reoffending.

## Methodology

The following section begins with an outline of the experiment's research design, describing how the RCT was conducted and the Checkpoint procedure, and provides a summary of some of the ethical considerations for the research.

Checkpoint is a 4 month adult offender deferred prosecution scheme which targets offenders entering the Criminal Justice System by providing an alternative to a criminal prosecution. Eligible offenders work with a Navigator on a bespoke 'contract to engage', in order to address their underlying causes of offending. On the basis of the Turning Point research study and the encouraging findings from the Checkpoint initial pilot implementation phase (Weir et al., 2019), a Randomised Controlled Trial (RCT) was registered and conducted in Durham between August 2016 and April 2018. The RCT was the preferred research design to test whether a deferred prosecution model supported

by desistance could reduce reoffending (Routledge, 2015). The experiment was designed to test the following hypothesis:

The Checkpoint contract offering a deferred prosecution to a target group of offenders, coupled to an offender-specific supported desistance programme, will reduce reoffending in comparison to conventional criminal justice procedures.

To address this hypothesis, the research questions are:

1. Is there a statistically meaningful difference in the prevalence of offending between the offenders who were assigned to the Checkpoint programme and those who were assigned to the normal criminal justice process?
2. Does the 'hazard' (or risk) of reoffending from the date of random assignment to the date of the first offence, during the follow-up period of 24-months, differ to a statistically significant extent when comparing the Checkpoint 'treatment' and normal criminal justice process 'control' cohorts?

Based on data collected as part of the RCT, two additional research questions are stated as follows:

3. Is there a statistically meaningful difference in the frequency of offending between the offenders who have successfully completed the Checkpoint programme and those who undergo the normal prosecution process?
4. Is there a difference in harm caused by offences before and after Checkpoint compared to the control group, based upon on the Office of National Statistics 'crime severity scores'?

Routledge (2015) describes how this hypothesis and associated research questions was tested. She describes the protocol for conducting a randomised control trial experiment in the form of 'Checkpoint', including the design and implementation, and summarises Phase One of Checkpoint, which is the setting up of the experimental environment and securing and testing the treatment.

### **Recruitment**

Subjects were recruited into the experiment following arrest at each of four custody suites across the Durham force area. Recruitment is a two-step process, the first utilising a forecasting model specifically developed for Checkpoint using Durham Constabulary data, and the second is the randomisation into the experiment.

Neyroud (2015a) highlights the benefits of applying an 'actuarial approach to risk assessment' and thus applying evidence-based practice to targeting treatment resources. A forecasting model, or Harm Assessment Risk Tool (HART), was developed specifically for Durham Constabulary using Durham custody and offender data (Barnes, 2016; Urwin, 2016). HART uses details of 104,000 custody events from 2008 to 2012 and data from 34 predictor variables – covering for example a person's age, gender and offending history, to create a prediction about each person. On the basis of a dataset over the 5-year period to build the model, and then a dataset for the following 2-year period to test it, the HART model provides a forecast as follows:

LOW – subject will not reoffend during the following 24 months;  
MODERATE – subject will reoffend during the following 24 months;  
HIGH – subject will commit a serious offence during the following 24 months.

This utilised the random forest procedure (a decision-making tool which builds multiple decision trees based on variables and amalgamates them together to get a more accurate and stable prediction) and has an ability to balance different types of errors, by using cost ratios decided by the organisation, in this case Durham Constabulary.

The forecast is predicated on predictor variables compiled and tested during the development phase of the model, each variable contributing to its predictive accuracy. Each forecast for each nominal and each custody episode is stored in the model database, using a unique identification number. Data for each of the predictor variables that were included in the final model design are collected as part of this, including the number of ‘votes’ for each of the low, moderate, and high risk levels assigned as part of the algorithm process. The risk level with the highest number of votes determines the result of the forecast.

The baseline characteristics analysis (Table 2) is based upon a selection of predictor variables that contribute most to the predictive accuracy of the model, and which are known in a criminological context to be particularly relevant to reoffending behaviour more generally. Hence, each unique forecast provides the data used to assess baseline characteristics across groups accordingly.

This model was used to identify how much of the offender population that is run through the forecasting model will be low, moderate, and high (noting that this means arrestees processed through custody). For the purposes of this research, Durham Constabulary’s Checkpoint programme targeted those offenders arrested and given a ‘moderate’ risk assessment. The rationale for this was that Checkpoint was designed primarily to support desistance from criminal behaviour(s) and offending, such that identifying any change (or statistical difference) in offending will be more achievable in a group that is likely to reoffend, particularly when compared to a similarly composed control group.

## *Process*

The process begins when a suspect is arrested, brought to police custody and booked into detention by a qualified custody sergeant and the officer in the case conducts an interview. The RCT recruitment process always starts by running a forecast, as outlined above, which uses the forecasting model embedded in the Client & Offender Management, Engagement Tracker (COMET) offender management application – a specific case management IT system developed specifically for Checkpoint. Once the forecast is run and the result is available, an eligibility check is made with the subject before the random assignment is made (Routledge, 2015). This process is carried out as follows:

- The randomisation application is triggered by the forecasting result, which enables the Custody Sergeant to enter the Police National Computer (PNC) unique identifier, the custody record reference number, and to go through the eligibility check question set with the person while at the custody desk.

Table 2. Baseline characteristics.

T-tests for Baseline Characteristics	Moderate RCT 'Intention to Treat' (ITT) analysis				Moderate RCT 'Per-Protocol' (PP) analysis							
	Control (ITT)	% or SD	Treatment (ITT)	% or SD	p values (two-tailed) vs Control	Cohen's d	Control PP	% or SD	Treatment PP	% or SD	p values (two-tailed) vs Control	Cohen's d
Male	176	76%	224	78%			168	75%	100	76%		
Female	57	24%	64	22%			56	25%	31	24%		
Cohort size	233	100%	288	100%			224		131			
<b>Age</b>												
Mean	32.3	10.1	33.9	9.4	p = 0.058	0.17	32.4	10.2	33.4	9.4	p = 0.363	0.10
Median	31		33.5				31.5		33			
<b>Age of Onset (mean)</b>												
Age at first offence	23.4	9.2	23.5	8.9	p = 0.850	0.02	23.5	9.3	23.5	9.3	p = 0.989	0.02
Age at first Violent offence	25.1	9.3	26.4	9.5	p = 0.179	0.14	25.1	9.4	26.5	9.8	p = 0.264	0.14
Age at first Property offence	22.6	8.1	23.3	8.7	p = 0.441	0.08	22.7	8.1	22.6	8.5	p = 0.936	0.01
<b>Presenting Offence count (mean)</b>												
No. of Presenting offences	1.5	1.4	1.5	1.3	p = 0.946	0.01	1.5	1.4	1.4	0.9	p = 0.417	0.09
<b>Prior Custody/Offence counts (mean)</b>												
No. of Prior Custody events	11.1	14.6	14.4	18.3	p = 0.023	0.20	10.7	13.5	11.0	14.9	p = 0.871	0.02
Prior offences count	16.5	23.1	21.2	28.0	p = 0.033	0.19	15.7	21.0	15.8	24.8	p = 0.963	0.01
Prior Violence offences count	3.0	3.8	3.6	4.6	p = 0.100	0.15	3.0	3.8	3.0	3.5	p = 0.952	0.01
Prior Property offences count	6.5	13.4	9.3	16.8	p = 0.038	0.18	6.3	12.4	6.4	13.5	p = 0.919	0.01
Prior Drug offences count	1.0	2.1	1.3	2.9	p = 0.148	0.12	0.9	1.8	0.9	2.6	p = 0.886	0.02

(continued)



**Table 2.** (continued)

T-tests for Baseline Characteristics	Moderate RCT 'Intention to Treat' (ITT) analysis					Moderate RCT 'Per-Protocol' (PP) analysis						
	Control (ITT)	% or SD	Treatment (ITT)	% or SD	p values (two-tailed) vs Control ITT	Cohen's d	Control (PP)	% or SD	Treatment (PP)	% or SD	p values (two-tailed) vs Control PP	Cohen's d
<b>Time since last offence (mean)</b>												
Time since most recent Custody	1.3	2.1	1.6	2.5	p = 0.147	0.13	1.3	2.1	1.7	2.6	p = 0.157	0.17
Time since most recent offence	1.3	2.1	1.6	2.5	p = 0.154	0.12	1.3	2.1	1.7	2.7	p = 0.160	0.17
Time since most recent Violent offence	2.8	3.3	3.6	4.1	p = 0.033	0.21	2.8	3.3	3.6	3.9	p = 0.094	0.22
Time since most recent Property offence	4.1	3.8	4.7	4.2	p = 0.125	0.16	4.2	3.8	5.5	4.4	p = 0.015	0.33
<b>Prior Intelligence entries count (mean)</b>	33.3	74.9	45.9	89.1	p = 0.080	0.15	32.9	75.3	31.0	67.6	p = 0.812	0.03
<b>Risk Forecast (mean of vote counts)</b>												
Low	99.2	52.6	93.5	54.5	p = 0.229	0.11	100.7	52.7	106.3	53.8	p = 0.344	0.10
Moderate	253.1	36.9	257.2	42.5	p = 0.244	0.10	251.9	36.5	250.1	39.0	p = 0.665	0.05
High	156.8	39.3	158.3	37.8	p = 0.644	0.04	156.4	39.3	152.6	36.6	p = 0.374	0.10

- If eligible, the randomisation app uses a computer-generated randomisation syntax to randomly assign the subject into one of two groups, displaying the result on screen as either ‘allocate to Checkpoint’ (treatment) or ‘normal criminal justice process’ (control).
- Depending upon the result of the above randomisation process, the random assignment provides the Custody Sergeant with the necessary information to finalise the disposal process for the detainee.

If randomly assigned to the ‘control’ group, the Custody Sergeant issues a sanction as per normal criminal justice processes accordingly.

If randomly assigned to the Checkpoint ‘treatment’ group, the Custody Sergeant provides a brief explanation of Checkpoint to the detainee, as well as a printed sheet with details, and, pending consent from the detainee, books an appointment with the Navigator team based at the respective custody site. In line with deterrence certainty and celerity principles, the meeting is scheduled at the earliest opportunity, usually within 24–72 hours and the offender becomes a Checkpoint subject. If the subject does not agree, s/he is processed as per traditional criminal justice processes.

### *The checkpoint intervention*

The Checkpoint intervention has previously been described by Weir et al. (2019) and so is described only in summary form here. The decision-making process for eligibility criteria and ‘included offences’ for the Checkpoint programme was made by the multi-agency Checkpoint Governance Board (Routledge, 2015; Weir et al., 2019). The key distinction between the pilot phase and the RCT was the addition of HART in the recruitment process to target those of relevant forecast risk.

Durham Constabulary developed the Checkpoint programme based upon Operation Turning Point (OTP) implemented by West Midlands police, with the Checkpoint research study being, for the most part, a replication of OTP. However, one of the key distinguishing factors between the approaches was in OTP using Police Officers as Offender Managers, while Checkpoint employed specialist Navigators, as outlined in the previous section. Moreover, four of the team of eight Navigators were seconded from the National Probation Service both for the pilot phase and the duration of the trial (Weir et al., 2019). Checkpoint also employed much of the methodology behind the Offender Group Reconviction Scale (OGRS) and the Offender Assessment System (OASys), both developed for the Prison and Probation Service in England and Wales (Debidin, 2009; Moore, 2015).

In the same way that the OGRS model seeks to predict reoffending over a 24-month period, the Durham forecasting model, HART, provides a forecast of the risk of re-arrest over an equivalent period. Furthermore, Checkpoint adopted a similar approach to OASys as part of the assessment process carried out by the Navigators. OASys is predicated on the Risk-Needs-Responsivity (RNR) model and is based on the notion that any treatment for offenders needs to be aligned to these three dimensions (Andrews et al., 1990a). This approach forms the basis of the Checkpoint assessment process to identify ‘criminogenic needs’ of those entering the Checkpoint programme.

Checkpoint clients complete an in-depth needs assessment, as part of their first meeting and interview with the Navigator, to determine why they have offended and what support they need to desist from further offending. The criminogenic needs identified, or ‘critical pathways’, are used to determine the needs of the individual and the actions which will form the interventions as part of the ‘contract’ to engage. The critical pathways used as part of the needs assessment are:

- Accommodation;
- Alcohol;
- Attitudes and Behaviour;
- Relationships (children and families);
- Drugs;
- Employment, Training and Education;
- Finance (budgets and debt);
- Mental and Physical health; and
- Exploitation (including sexual exploitation and Modern-Day Slavery & Human Trafficking).

In assessing the critical pathways of offenders, several assessment tools are already in use by other health organisations, for example the Alcohol Use Disorders Identification Test (AUDIT) (NHS, 2015) and these are used by the Navigators. The relevant organisations were consulted on to determine what needs assessment tools they have based on the needs above. This has then been collated into a needs assessment form that is used by the navigators to complete during the first interview with offenders. Subjects complete a self-assessment survey measuring perception of current life problems (McDougall et al., 2009; Sadlier, 2010) and are asked to rate on a scale of 1–10 how much they believe this area of need is impacting on their offending, and to identify which three contributed the most.

From this assessment, a Navigator creates a bespoke, individually tailored 4-month ‘contract to engage’ with up to five conditions (Routledge, 2015; Weir et al., 2019). It includes a mandatory condition, an agreement to not reoffend and, where applicable, an agreement to meet the victim to put right the harm caused (Restorative Approach). The other conditions will include interventions around the reasons for offending, or critical pathways. There is also an option for the offender to wear a GPS tag and/or complete between 18 and 36 hours of voluntary work to give something back to their community and to develop their personal skills, but it is important that this is seen as an opportunity to rehabilitate rather than penalise.

The offence is classed as a deferred prosecution, which can be invoked at any point should the subject breach the contract. If the offender complies with their contract, they will face no further sanctions from the Criminal Justice System. If they fail, the prosecution will be invoked and they will attend court (subject to the Crown Prosecution Service (CPS) Charging Standards if required), or another disposal commensurate with the offence.

Lastly, an important aspect of the intervention is the work of the Navigator, both in terms of mentoring the offender through the contract and monitoring progress with

critical pathway referrals. Navigators can use the COMET system to administer their work with each client, including all details of the initial assessment and activity log. Data collected in COMET in this way can be used to assess the ‘treatment integrity’ of the Checkpoint intervention, specifically in terms of (Durlak and DuPre, 2008):

- I. Fidelity – the extent to which the programme corresponds to the originally intended programme;
- II. Dosage – how much of the programme is delivered;
- III. Quality – how well the different programme components have been delivered.

### *Outcome measures*

If the hypothesis tested in Checkpoint is to withstand scrutiny and change policy, not only must the research design be sufficiently robust and maintain statistical power but also the measurement of the outcomes from the RCT must be reliable and valid. The intention of this study is to establish if reoffending and harm is different within the treatment versus control groups of Checkpoint’s RCT. Chi-square tests were conducted to consider the expected and observed proportions of offending in the treatment and control groups, and t-tests were used to establish to what extent the two groups differ in terms of both frequency and harm of offending (24-months pre- and post-intervention). Ancillary analysis was also conducted using the Cox Proportional-Hazards Regression model to incorporate the time to reoffending in the 24-months following random assignment, identifying the hazard ratio of reoffending between the treatment and control groups. Lastly, given any degradation in experiment, samples based upon varying levels of experiment integrity were identified for the analysis.

*Reoffending.* In terms of reoffending, the analytical strategy used here applies event history and survival analysis techniques to quantitative data sets in order to address the key research question of whether Checkpoint can sustainably reduce reoffending in comparison to normal prosecution processes (Brennan, 2015). Police National Computer (PNC) checks for both ‘police disposals’ and ‘court sanctions’ will be employed to provide a statistic for reoffending to determine frequency, prevalence, time-to-failure. The PNC checks will be conducted on the basis of offences that were committed during the 2-year follow-up period (from the date of random assignment). This differs to the Ministry of Justice (MoJ) definition of proven reoffending which looks at reoffending in a 1-year follow-up, with a further 6 months to allow for court processes. This means that the definition for proven reoffending employed by the MoJ still holds, in terms of verification that a police disposal or court sanction has taken place, but extends the monitoring period from the date of random assignment to the date when a reoffending episode takes place from 12-months to 24-months.

*Harm.* In terms of harm, the Office of National Statistics Crime Severity Score (ONS CSS, 2017) weights offences based on sentencing information sourced from the MoJ. Custodial sentences, community orders and fines have been included in the construction of the weights, with offence weight values calculated as the proportion of offenders

receiving the type of punishment multiplied by the average sentence length in days (or prison length equivalent in terms of orders and fines). While the Cambridge Harm Index takes the minimum sentencing tariff for each offence, and weights each offence accordingly (Sherman et al., 2016), the ONS Severity Score takes the average (mean) tariff. The ONS methodology takes this a step further by weighting this average tariff score according to the proportion of offenders who receive custodial sentences, community orders, and fines. The decision was taken to use the ONS CSS rather than the Cambridge Crime Harm Index, and that harm would be measured by comparing the sum of crime severity scores for all offences in the 24-months prior to random assignment with the sum of CSS for all offences in the 24-months following random assignment.

### *Ethical considerations*

Punch (1986) and Miles and Huberman (1994) raise a number of ethical issues in relation to conducting police research.

The random allocation of potentially prosecutable cases does raise important ethical considerations for the Criminal Justice System. One of these is to ensure that there is no breach as to the fairness of the process. If randomly assigned to the Checkpoint 'treatment' group, the treatment should not provide a disproportionately tough set of conditions, cause them harm or worse consequences than if they were otherwise assigned to the control group (Edwards et al., 1998; Torgerson and Torgerson, 2008).

In satisfying the conditions as outlined above, the recruitment process should not start until a clear decision for a positive disposal has been taken by the Custody Sergeant. Hence, the alternative to participation in the RCT could be prosecution leading to a criminal conviction and an appropriate sanction. If randomly assigned to the Checkpoint treatment programme, the prosecution is deferred. The contract conditions have been designed by drawing on the theoretical evidence discussed previously, so as to ensure that conditions offered are supported by the best evidence available and support the offender positively to reduce reoffending and improve their life chances.

Secondly, if an offender does not admit to the offence, this will not of itself lead to mandatory exclusion from the experiment. If offenders are deemed suitable for prosecution because they satisfy the sufficiency of evidence test, they are also eligible (subject to passing through the eligibility check as part of the recruitment process). There is also the issue of the offender's consent to participate in the experiment. The Checkpoint process enables consent to be sought in two stages:

- I. Initial agreement to consider the experiment and return voluntarily to the police station for their first Navigator appointment (but with the threat that failure to attend could trigger prosecution);
- II. Fully informed consent, following a structured interview with the Navigator.

Lastly, there needs to be a consideration of the victim's views about the decision to place an offender in the experiment. According to the Code of Practice for Victims of Crime (2006), the police and CPS have an obligation to ensure that the needs of vulnerable victims, or victims with special needs, are taken into account and that victims are

kept informed about significant events or changes in the investigation, such as whether the offender will be cautioned, given the opportunity to participate in diversion scheme or charged (noting that the decision to divert an offender ultimately remains with the police). The Checkpoint process incorporates this by enabling victims to be kept up to date with the offender's progress if they so wish and, if appropriate, gives the victim an opportunity to have a say when setting out the conditions of the contract with the offender, for example, by way of including restorative approaches.

## **Findings**

Intention-to-Treat (ITT) analysis is used to test the effectiveness of the intervention, on the basis that this preserves the integrity of the experiment and the efficacy of the analysis, in terms of baseline characteristics. Table 2 offers a perspective on the baseline characteristics such as gender, age, previous offences, risk. The analysis of baseline characteristics across both ITT and PP cohorts found no meaningful differences.

However, given the constraints of criminal justice settings, a number of eligible subjects declined to participate following random assignment, with a further group unable to receive their assigned intervention due to recruitment errors, with both of these factors having a disproportionate impact on the Checkpoint 'treatment' cohort.

Looking first at the control group, of the 233 randomly assigned to this arm of the trial, 224 or 96% received the designated control intervention, representing a high level of experiment integrity. However, of the 288 randomly assigned to the Checkpoint treatment group, 131 or 45% received the designated intervention, i.e. entered into the Checkpoint contract.

In terms of the ITT analysis, this includes all those randomly assigned to each of the treatment and control groups, irrespective of whether they actually received their assigned intervention. When looking at the PP analysis, this incorporates only the 96% of the controls that actually received their assigned intervention and, likewise, only the 45% of the treatment group that actually received the Checkpoint intervention.

Participant flow is illustrated in Figure 1.

### ***Research question 1 – Prevalence of reoffending***

Using ITT analysis, results showed that the prevalence of reoffending in the Checkpoint treatment group ( $M = 0.42$ ,  $SD = 0.5$ ) was slightly less than in the control group ( $M = 0.46$ ,  $SD = 0.5$ ), but an independent-samples t-test didn't find a significant difference,  $t(433) = 0.65$ ,  $p = 0.52$ .

Using PP analysis, an independent t-test found that there was a significant difference between the control ( $M = 0.47$ ,  $SD = 0.5$ ) and treatment ( $M = 0.37$ ,  $SD = 0.5$ ) cohorts that did actually receive their intended intervention,  $t(283) = 1.9$ ,  $p = 0.055$ , with a small effect size, Cohen's  $d = 0.21$  (Cohen, 1998).

Figure 2 illustrates these results.

Table 3 summarises the cohort results given in the above chart.

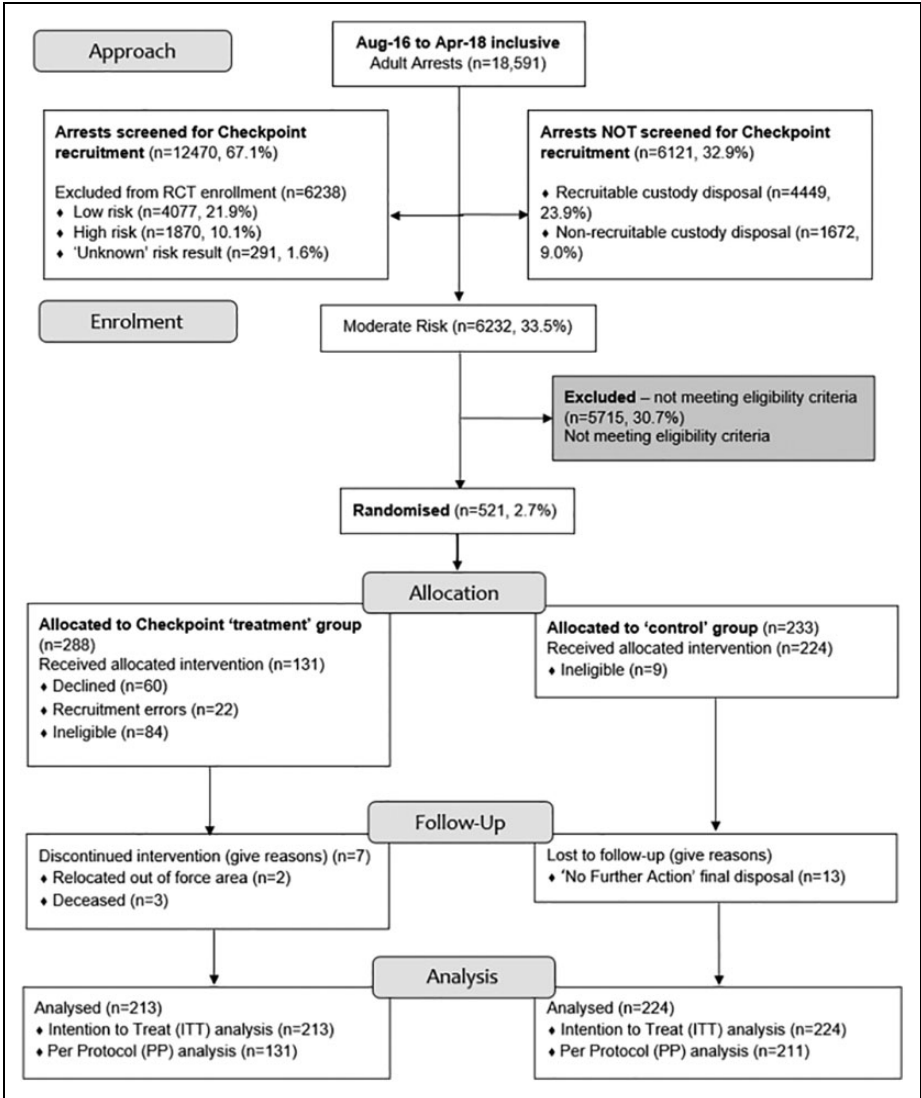


Figure 1. CONSORT flow diagram.

### Research question 2 – Time to reoffending

As an extension to the analysis for prevalence, Cox Regression survival analysis provides an estimate for the hazard ratio, or relative risk, of reoffending in the ITT cohorts: 0.91 (95% confidence interval or CI for the hazard ratio is 0.69–1.21). This represents a 10% decrease in risk on average for those assigned to Checkpoint treatment when compared to the control ITT cohort, although the result is not statistically significant ( $p = 0.51$ ).

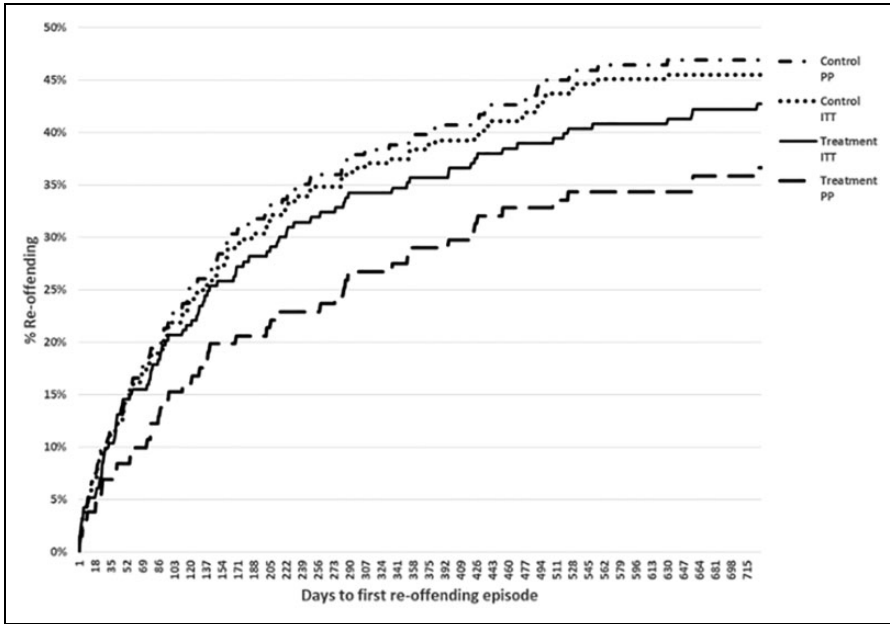


Figure 2. Prevalence of reoffending for RCT cohorts.

Table 3. Prevalence of reoffending for RCT cohorts.

Prevalence of Reoffending	Control ITT	Checkpoint ITT	Control PP	Checkpoint PP
Cohort	224	213	211	131
Reoffending Count	102	91	99	48
Reoffending %	45.2%	42.7%	46.9%	36.6%

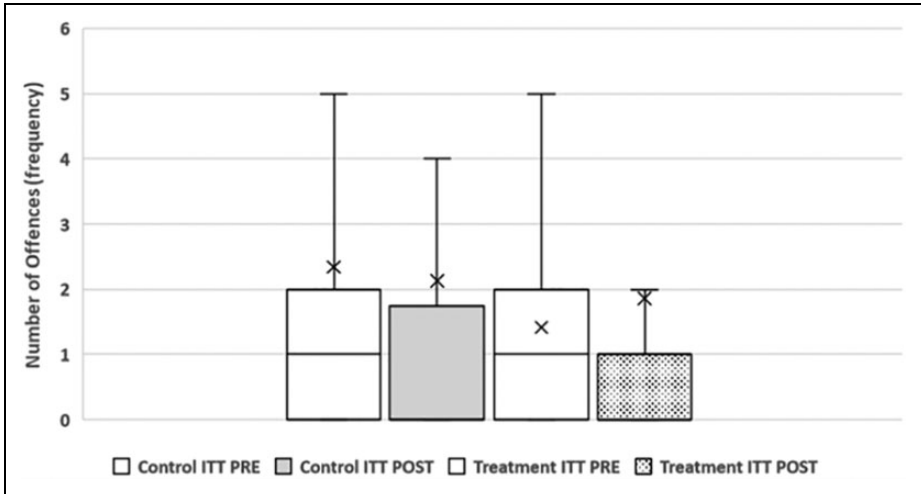
However, when comparing the PP cohorts, the hazard ratio is 0.7 (95% CI 0.49–0.99) and is significant ( $p = 0.04$ ), representing a 30% decrease in the risk of reoffending with those who actually received the Checkpoint treatment compared to those in the control that actually received the assigned intervention.

### Research question 3 – Frequency of reoffending

This section considers how the number of offences in the 24-months prior to random assignment compares to the number of offences in the 24-months following random assignment, both assessed on the basis of whether the offence received a police disposal or court sanction, i.e. proven reoffending.

*ITT cohorts.* Looking first at the ITT cohorts, the controls saw a slight reduction, on average, from pre-intervention ( $M_{pre} = 2.34$ ,  $SD_{pre} = 5$ ) to post-intervention ( $M_{post} = 2.13$ ,





**Figure 3.** Box plots for frequency of reoffending for ITT cohorts.

$SD_{post} = 5.2$ ), while those assigned to treatment saw a small increase ( $M_{pre} = 1.41$ ,  $SD_{pre} = 2.6$ ;  $M_{post} = 1.9$ ,  $SD_{post} = 4.8$ ).

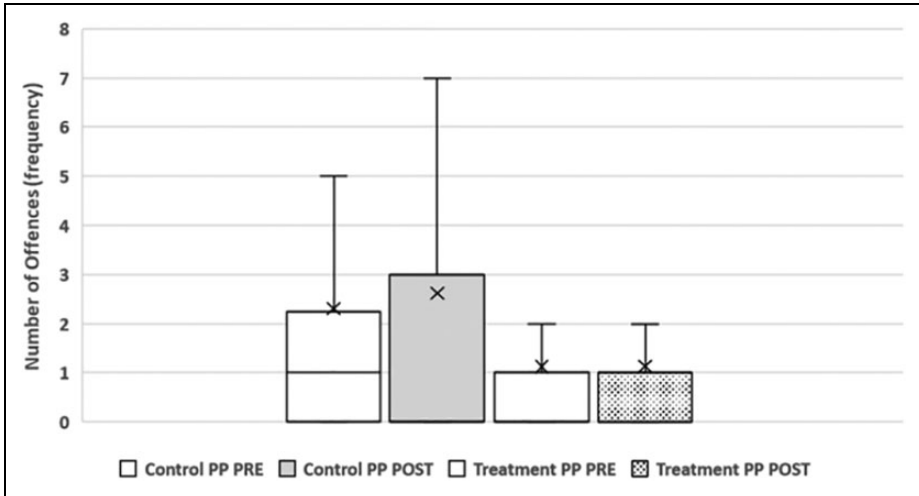
When looking at the median values and interquartile ranges (IQR), the control ITT median is 1 offence for pre-intervention frequency (Interquartile range or IQR 0–2 offences) and zero offences for post-intervention frequency (IQR 0–1.75). For the Checkpoint ITT cohort, median value remained at 1 offence for both pre- and post-intervention (IQR pre 0–2 & post 0–1). These results are illustrated in the box plots chart in Figure 3.

To contextualise how the box plot represents the data, half or 50% of the cohort is shown by the box component or interquartile range, e.g. half of the subjects in the control ITT cohort had between 0 and 2 offences pre-intervention, and this is the same for the treatment ITT cohort. The cross in the figure represents the mean value for the number of offences.

**PP cohorts.** When considering the PP cohorts, the controls increased the frequency of offending from pre-intervention ( $M_{pre} = 2.31$ ,  $SD_{pre} = 4.8$ ) to post-intervention ( $M_{post} = 2.61$ ,  $SD_{post} = 5.6$ ). Those who received the Checkpoint treatment saw no change in the level of offending both pre- and post-intervention ( $M_{pre} = 1.12$ ,  $SD_{pre} = 2.2$ ;  $M_{post} = 1.13$ ,  $SD_{post} = 3.1$ ).

The control PP median is 1 offence for pre-intervention frequency (IQR 0–2.25 offences) and 0 for post-intervention frequency (IQR 0–3). For the Checkpoint PP cohort, median value remained at zero offences for both pre- and post-intervention (IQR pre 0–1 & post 0–1). These results are illustrated in the box plots chart in Figure 4.

However, independent t-tests found neither the ITT nor PP cohort differences to be statistically significant ( $p > 0.23$ ), with all Cohen's d values suggesting small effect sizes ( $d < 0.2$ ).



**Figure 4.** Box plots for frequency of reoffending for PP cohorts.

#### Research question 4 – Harm

As with the analysis provided for the frequency of reoffending, the harm attributed to each offence using the ONS Crime Severity Score (CSS) was compared in the 24-months period pre- and post-intervention. Using the paired samples t-test:

- The control ITT group saw an increase in average CSS values for each offence ( $M_{pre} = 30$ ,  $SD_{pre} = 69.9$ ;  $M_{post} = 37$ ,  $SD_{post} = 90.3$ ),  $t(1004) = 1.53$ ,  $p = 0.13$ , Cohen's  $d = 0.09$ , although the median CSS value remained at 12 from pre- to post-intervention, with interquartile range (IQR) pre: 8 to 14 and post: 12 to 15.
- CSS values in the treatment ITT group also increase from pre- to post-intervention ( $M_{pre} = 24$ ,  $SD_{pre} = 57$ ;  $M_{post} = 69$ ,  $SD_{post} = 420.4$ ),  $t(474) = 2.22$ ,  $p = 0.03$ , Cohen's  $d = 0.15$ , with a median value of 12 remaining consistent in both periods (IQR pre: 8 to 15, post: 12 to 15).

Median and IQR are represented using box plots in Figure 5, with outlier points excluded from the chart for clarity but not from the underlying statistical analysis.

Using box plots to represent the data, the mean values are statistical outliers, which indicates the disproportionate impact that the higher CSS values can have on the analysis. Findings predicated entirely on differences between statistical outliers, in the true sense of the word, presents a key issue for interpretation, since outliers are unrepresentative of the majority (just over 99%) of the data. The treatment ITT result was impacted by four offences with very high CSS values, two offences committed by one individual, and one offence each by two other individuals. The pattern of these results is largely replicated in the PP cohorts and so is not presented here. This indicates the sensitivities in analysing harm and the impact that outliers can have on the overall harm score for the cohort.

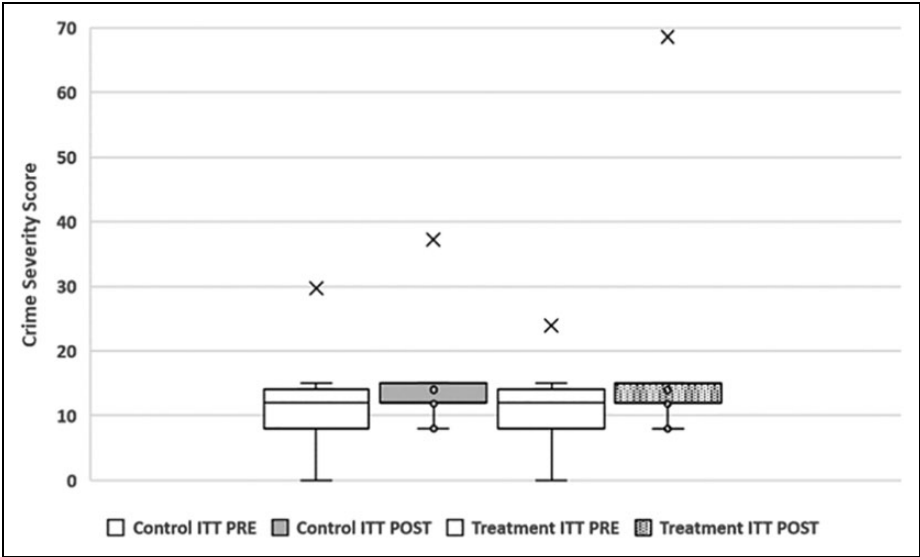


Figure 5. Box plots for harm pre and post-intervention for the ITT cohort.

### Discussion

The evaluation provides a reasonable degree of confirmation that the Checkpoint intervention can reduce reoffending in comparison to traditional criminal justice procedures, thus addressing the overarching hypothesis set out in the research study. The findings from this randomised control trial offer evidence that a deferred prosecution model based on desistance, deterrence, and supported critical pathways may reduce reoffending compared with traditional criminal justice processes. These findings are in line with the evidence that underpins the principles of offender desistance and deterrence, as outlined in the literature review. However, there are complex and complicated findings that underpin this result, which indicate some caution should be applied.

#### Research question 1 – Prevalence of reoffending

While both the ITT and PP cohorts demonstrated a reduction in reoffending in favour of the Checkpoint intervention, only the PP analysis provided a statistically significant test result, a 10.3% reduction in the prevalence of reoffending (46.9% of the control cohort reoffended compared to 36.6% of the Checkpoint cohort).

#### Research question 2 – Time to reoffending

Results indicate a 30% reduction over time, in the hazard or risk of reoffending during the 24-month follow-up study period. The hazard ratios suggested by the survival analysis with the ITT cohorts indicates a potential reduction in offender caseload. On average, there is a 10% reduction in the hazard, or risk over time, of reoffending over the

24-month follow-up study period. To contextualise these results, for every 100 offenders there will be 10 less who reoffend if they were given Checkpoint compared to normal criminal justice outcomes. Taking all the hazard ratio results for each ITT sample by experiment integrity, then this reduction could range from just 2 less to as many as 33 less reoffenders following Checkpoint for every 100 control reoffenders. For the per-protocol cohort, if the confidence intervals in the statistically significant PP cohorts are taken into account, this could range from 5 less to as many as 4 less Checkpoint reoffenders for every 100 reoffenders who were given a normal criminal justice sanction.

The Checkpoint research study targeted a ‘moderate’ risk cohort of arrestees, which corresponded to about half of all arrestees processed through the custody suites for non-administrative disposals. In targeting moderate risk offenders defined in this way, this introduced a degree of heterogeneity (or diversity) into the cohort which is best exemplified by baseline characteristics such as the number of prior custody events or prior offences count (see Table 2). For example, the mean count of prior arrest offences for the PP control and treatment groups was 15.7 and 15.8 respectively. The rationale for targeting a higher risk level of offender was to determine whether an experimental intervention could reduce reoffending on the basis that it would be easier to identify any such difference with a cohort that is expected to reoffend. In contrast, Operation Turning Point specifically targeted a lower risk cohort, excluding those with more than one previous conviction, those who would probably receive a custodial sentence, or those with a serious presenting offence (Neyroud, 2015b). Of these three possible reasons for exclusion, only a ‘serious presenting offence’ would prevent entry into the Checkpoint study. The impact of targeting a higher risk cohort is most readily apparent in the findings for the additional research questions concerning frequency of offending and harm. This seems to be in agreement with Parker and Grasmick (1979) who suggested that the threat of legal sanctions is only successful for a proportion of the population as some individuals are beyond correction or reform. It may well be that there is a substantive difference between those nominals forecast at the higher end of the ‘moderate’ band, in that they are generally entrenched prolific offenders, as distinct from the lower ‘moderate’ band or those lower risk offenders targeted by Operation Turning Point.

### *Research question 3 – Frequency of reoffending*

The rationale for considering the frequency of offending pre- and post-intervention was to establish whether any deceleration in offending had taken place. The difference in average or mean values for the frequency of offending across the ITT and PP cohorts in the 24-months before and after random assignment did not reach statistical significance. Therefore, it may be helpful instead to consider what the box plot analysis tells us.

In terms of the ITT cohorts, the control group saw half of the subjects in the group with between 0 and 2 offences in the 24-months pre-intervention, and this IQR changed to between 0 and 1.75 offences in the 24-months post-intervention. For those assigned to the Checkpoint ITT ‘treatment’ group, results were exactly the same as the controls pre-intervention (IQR 0–2) but between 0 and 1 offence post-intervention. Moving on to the PP cohorts, a different picture emerges, with the controls this time increasing their IQR

from 0 to 2.25 offences pre-intervention to 0 to 3 offences post-intervention. However, the Checkpoint PP cohort maintained the same IQR before and after of 0 to 1 offences. To offer some context to these results, Durham Constabulary has generally dealt with approximately 4000–4500 subjects each year (over the past 3 calendar years) by way of arrest and entry into a criminal justice process. For clarity, around 4500–5000 subjects each year and over the same period were arrested but released without bail or no further action taken. Taking only those who would be risk assessed as ‘moderate’, approximately 2250 subjects each year, then a difference in IQR of 0 to 1.75 and 0 to 1 post-intervention is indicative of the potential for less victimisation and fewer victims.

With ITT controls displaying a slight decrease in frequency of offending post-intervention, but an increase when only considering PP controls, this highlights the impact that a group of just 13 nominals can have who were randomly assigned to the control group but were eventually given a ‘no further action’ final disposal. Similarly, the Checkpoint ITT saw a small increase in frequency post-intervention, but with no change before and after when looking at the Checkpoint PP cohort. For the Checkpoint cohorts, this highlights the impact of the 82 nominals who were randomly assigned to receive the Checkpoint programme but didn’t. On balance, even accounting for the surprising impact of just 13 subjects in the control cohorts, the overall result for achieving any deceleration in offending from either intervention is inconclusive.

#### *Research question 4 – Harm*

When looking at harm pre- and post-intervention, the context for these results is important in that the mean or average crime severity score (CSS) values across both ITT and PP cohorts are well outside their respective interquartile range (or where half of the CSS values lie). The median and IQR values of harm generally remained stable from pre- to post-intervention across both the control and treatment groups. In contrast, the Turning Point experiment reported 8.2% lower average harm scores, using the Cambridge Crime Harm Index (Sherman et al., 2016), and a larger reduction of 36% when just looking at violent offenders (Neyroud et al., 2015). A possible explanation for what happened in the Checkpoint experiment in relation to the harm results may lie in the higher, although not high, risk cohort that the Checkpoint study targeted. On average, approximately 30% of subjects in each of the Checkpoint study treatment and control groups were responsible for 88% of the total number of offences. While Turning Point restricted entry to those with no more than one conviction, the average conviction count pre-intervention across both arms of the Checkpoint experiment was approximately 4. The proposition here is that a minority of individuals in the experiment, irrespective of which arm they were assigned to, were sufficient to skew any benefit accrued in the majority of the study cohort. It is beyond the remit of this paper to unpack the complexities inherent in the analysis of harm using police data (c.f. Ashby, 2018; Higgins, 2017) but, again, the context can be helpful here.

This also called into question whether the integrity of the Checkpoint treatment or intervention, was robust. Treatment integrity carries enormous importance when testing any hypothesis in an experiment (Neyroud and Slothower, 2013). The internal validity of the experiment is ensured providing the RCT is implemented in line with the protocol,

and there is fidelity to the treatment and control groups, and the treatment processes. There was sufficient tracking and monitoring of data and information to support the assertion that the interventions, whether for control or treatment arms, were delivered consistently according to protocol standards to suggest that the treatment integrity was high. In summary:

- 78% of the Checkpoint contracts were successfully completed;
- On average, there were 4.4 critical pathways per contract;
- Navigators typically spent just over 8 hours 30 minutes with clients and associated client record maintenance; and
- In terms of counselling approaches, Navigators generally delivered a core of four or five distinct interventions to their clients, in combination with partner agency interventions.

Tables and charts are provided in the appendices.

However and as mentioned previously, the integrity of the experiment was less certain due to the nearly 40% loss of subjects in the treatment arm, mainly due to those who declined to participate or as a result of recruitment errors. While the control arm of the RCT achieved 94.2% compliance with assignment and receipt of the intervention, the impact of declines and recruitment errors on the Checkpoint treatment arm reduced compliance and receipt to 61.5%. Neyroud posits that when the compliance with the assigned intervention falls below a 60% threshold, this '[implies] that credible results are less likely to be obtained' (2017: 70). The baseline characteristics of those who either 'declined' or were 'recruited in error' into the experiment were compared to the treatment cohort as a whole to establish to what extent bias would be introduced into the analysis, particularly in relation to per-protocol. The reason why 80 subjects (28% of the treatment cohort) were recruited in error is because they were found to be ineligible after they were randomly assigned to the treatment group, mainly due to existing licence or supervision arrangements with probationary services. The baseline characteristics of this group reflect this in that they have more recently been in police custody or offended than the rest of the cohort. The process of randomisation means that they will most likely be a similar proportion of ineligible subjects assigned to the 'control' group. However, excluding only those 'recruitment errors' from the treatment group will introduce bias into the per-protocol analysis.

The Checkpoint study was the first time Durham Constabulary utilised a randomised control trial, and all the staff and officers involved were new to what is arguably the most advanced of research designs. The intervention, as formulated by Offender Desistance Policing, has only otherwise been tested with Turning Point by West Midlands Police; all of which suggests that, in every respect, Durham has met the conditions for what Neyroud defines as 'Novice theory' (2017). As outlined above, a key limitation for such experiments is the recruitment of sufficient subjects into the experiment given that it is taking place in 'real' custody conditions and the ability of custody sergeants to maintain a high level of compliance in processing subjects into the forecasting tool and when randomising subjects into the experiment.

Despite the inexperience of the organisation and its people in conducting policing research at this level, the Checkpoint research study has provided a powerful insight across a spectrum of offender management processes. Two examples of this were in understanding the intensity of the intervention should mirror the level of forecast risk and the importance of the relationship between the Navigator and offender. With respect to the first, there is potential to 'dial' the Checkpoint intervention to more accurately address the risk of reoffending posed by the individual, which could have a significant impact on how such programmes are resourced. Further, as a result of the bespoke nature of the intervention, Navigators can accommodate refinements in the programme that can compensate for the potential anomalies inherent when employing algorithmic forecasting techniques.

The findings presented here are the most central to the research questions originally posed but the study has revealed some important features of Checkpoint that would benefit from further research. For example, exploratory analysis has shown that theft offenders are more likely to reduce or stop their reoffending than violent offenders, and results from voluntary attenders who participate in the Checkpoint programme suggest that it can be highly effective compared to those who were arrested and participated in Checkpoint. Similarly, the differential effect of different contracted treatment interventions could benefit from further research. While any analysis may be prone to volatility, exploratory analysis has shown that across the contracted treatment cohort, a similar proportion had completed or failed across all pathways, except for alcohol and drugs. For alcohol, two-thirds of those contracted with this pathway successfully completed, and the other third failed; whereas for drugs, the opposite was found, with only one third successfully completing with this pathway, but the remaining two-thirds failing. On the basis of one RCT research study, the 4-month contract appears to be able to address those with less serious alcohol problems but was found to be generally insufficient to address those with drug problems. Looking at the literature, there are parallels in that for drug rehabilitation to achieve a sustainable impact, it requires a long-term programme and commitment from all parties, but often with relapses along the way, and so Checkpoint may only be viable with those who use such substances for entirely recreational use and/or on an infrequent basis. The use of an exit interview with those who successfully complete the Checkpoint programme lends itself to further qualitative research with participants. Lastly, given that this study focussed on a moderate risk cohort of offenders, further studies could identify to what extent Checkpoint is generalisable to a wider offender population.

## **Conclusion**

In reviewing cited research from the MOJ review of out-of-court disposals (CJJI, 2011), it became clear that there was a demand for high quality, empirical evidence to determine whether changes to OOC policy needed to be made. The sparsity of research into deterrence and desistance work with offenders, particularly pre-conviction studies in the UK, highlighted a gap in the evidence base that the Checkpoint randomised control trial aimed to fill. OOCs can account for between a quarter and a third of the people processed by the police. Therefore, improvements in practice in this area can bring

significant benefits not only for policing but with the potential for less demand on the courts and, ultimately, prisons. At a time when OOCs are being reviewed by the Ministry of Justice, it is pertinent that the evaluation findings are considered alongside any cross-government national reviews or strategies.

The hypothesis as originally set out was to establish whether the Checkpoint intervention could reduce reoffending and the research as set out in this paper suggests that it can. A reduction in the overall proportion that go on to reoffend in a 24-month period following random assignment can extend to just over 10% when comparing Checkpoint participants to controls, but this may be too narrow a perspective for this evaluation. It is not just the proportion of offenders that is important here, but the amount of offending that they are likely to perpetrate. Our study suggests that Checkpoint can reduce the risk of offending over a significant follow-up period and that may have far-reaching benefits for how policing resources are used. Moreover, while the hypothesis was a crucial motive for the research, there was an overarching ambition for the Checkpoint project, both in terms of working with offenders to address their criminogenic needs and in providing the opportunity to avoid further criminalisation. Putting aside arguments about what the remit of the police is or should be, working with offenders rather than simply processing them was a key tenet for the project and in the operationalisation of the programme.

While the evaluation has provided some clear evidence of how effective the Checkpoint intervention can be, some caution should be applied to the results in relation to the integrity of the experiment. The Checkpoint RCT highlights a key issue for experimental criminology in policing research, and the main limitation for the research study overall, namely the integrity of the experiment. For clarity, the integrity of the respective interventions applied to both the treatment and control groups met all required standards, but it was the loss of a large group of subjects from the Checkpoint treatment arm that was an issue. The issue centres on the imbalance introduced by those who didn't receive the intervention that was assigned to them. Under ITT analysis, the Checkpoint arm is diluted; whereas under PP analysis, an imbalance is introduced by excluding a sizeable group of subjects from one cohort and not the other. Such an imbalance introduces the prospect of 'selection bias' into the analysis, which will undermine the benefits of the RCT research design. While the analysis of baseline characteristics across both ITT and PP cohorts found no meaningful differences, this doesn't remove the risk of selection bias. Whist experience and knowledge suggest that the baseline characteristics presented in this paper are amongst the most important, there may be many other baseline characteristics for which data cannot be, or have not been, collected. The importance of this has not been ignored and is the subject of on-going research. With this very much in mind, utilising research designs such as randomised controlled trials enable analysis of an entirely different order to that provided by research normally applied in police settings. The call for further research is not just restricted to areas peripheral to the main Checkpoint programme but is equally important for the replication of research studies. If policy is to be guided by the evidence in a meaningful way, this will require meta-analysis developed on the basis of research in many if not most police forces, rather than from initiatives spear-headed by a select few.




## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Stephanie Kilili  <https://orcid.org/0000-0001-5738-519X>

## Supplemental material

Supplemental material for this article is available online.

## References

- Andrews DA, Bonta J and Hoge RD (1990a) Classification for effective rehabilitation: rediscovering psychology. *Criminal Justice and Behavior* 17(1): 19–52.
- Andrews DA, Zinger I, Hoge RD, et al. (1990b) Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology* 28: 369–404.
- Ashby MP (2018) Comparing methods for measuring crime harm/severity. *Policing: A Journal of Policy and Practice* 12(4): 439–454.
- Barnes GC (2016) *Focusing Police with Algorithms in Durham [Evidence-Based Policing Conference held at the Faculty of Law]*. Cambridge: University of Cambridge, 13 July.
- Beccaria C (1767) *An Essay on Crimes and Punishments*. London: J. Almon, opposite Burlington-House.
- Blakeborough L and Pierpoint H (2007) *Conditional Cautions: An Examination of the Early Implementation of the Scheme*. London: Ministry of Justice.
- Brennan I (2015) Centre for Criminology and Criminal Justice report: An experimental evaluation of an adult female triage pilot project for Humberside Police. National Police Library. Available at: <http://library.college.police.uk/docs/Police-Female-Triage-Report-Hull-University-2015.pdf> (accessed 18 October 2015).
- Cohen J (1998) *Statistical Power Analysis for the Behavioural Sciences* (2nd edn). Hillsdale, NJ: Erlbaum.
- Criminal Justice Joint Inspection (CJJI) (2011) Exercising discretion: the Gateway to Justice, 1–38. Available at: [https://www.justiceinspectorates.gov.uk/cjji/wpcontent/uploads/sites/2/2014/04/CJL\\_20110609.pdf](https://www.justiceinspectorates.gov.uk/cjji/wpcontent/uploads/sites/2/2014/04/CJL_20110609.pdf) (accessed 20 October 2015).
- Debidin M (2009) *A Compendium of Research and Analysis on the Offender Assessment System (OASys) 2006-2009*. Wellington: Ministry of Justice. Available at: <http://nomsintranet.org.uk/roh/official-documents/Debidin%20Compendium%20of%20OASys%20research.pdf> (accessed 20th April 2020).
- Durlak JA and DuPre EP (2008) Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology* 41(3–4): 327.
- Edwards SJL, Lilford RJ, Jackson JC, et al. (1998) Ethical issues in the design and conduct of randomised controlled trials. *Health Technology Assessment* 2(15): 1–146.

- Farrall S and Calverley A (2005) *Understanding Desistance From Crime*. London: McGraw-Hill Education (UK).
- Farrington DP (2003) Key results from the first forty years of the Cambridge study in delinquent development. In: Thornberry TP and Krohn MD (eds) *Taking Stock of Delinquency: An Overview of Findings from Contemporary Longitudinal Studies*. New York, NY: Kluwer Academic/Plenum, pp. 137–183.
- Gendreau P and Ross RR (1981) Correction potency: treatment and deterrence on trial. In: Roesch R and Corrado RR (eds) *Evaluation and Criminal Justice Policy*. Beverly Hills, CA: Sage, pp. 29–57.
- Gendreau P, Little T and Goggin C (1996) A meta-analysis of the predictors of adult offender recidivism: What works!. *Criminology* 34(4): 575–608.
- Giordano PC, Cernovich SA and Rudolph JL (2002) Gender crime and desistance: Towards a theory of cognitive transformation. *American Journal of Sociology* 107: 990–1064.
- Harvey E, Shakeshaft A, Hetherington K, et al. (2007) The efficacy of diversion and aftercare for adult drug-involved offenders: a summary and methodological review of the outcome literature. *Drug and Alcohol Review* 26: 379–387.
- Higgins A (2017) *Mixed Signals for Police Improvement: The Value of your Crime Severity Score May Go Up As Well As Down*. Perspectives on Policing, Paper 1, The Police Foundation, London, UK. Available at: [http://www.police-foundation.org.uk/2017/wp-content/uploads/2017/08/perspectives\\_on\\_policing\\_paper\\_1.pdf](http://www.police-foundation.org.uk/2017/wp-content/uploads/2017/08/perspectives_on_policing_paper_1.pdf) (accessed 20 April 2020).
- Hobbes T (1651 [1994]) *Leviathan*. In: Curley E (ed.) *Leviathan, With Selected Variants From the Latin Edition of 1668*, Indianapolis, IL: Hackett.
- Kennedy D, Kleiman M and Braga A (2017) Beyond deterrence: strategies of focus and fairness. In: Tilley N and Sidebottom A. (eds) *Handbook of Crime Prevention and Community Safety* (2nd edn). London: Routledge.
- Klepper S and Nagin D (1989) The deterrent effect of perceived certainty and severity of punishment revisited. *Criminology* 27(4): 721–746.
- McDougall C, Perry A, Clabour J, et al. (2009) *Evaluation of HM Prison Service Enhanced Thinking Skills Programme. Report on the outcomes from a Randomised Control Trial*. Ministry of Justice Research Series 3/09. London, UK: Ministry of Justice. Available at: <http://www.justice.gov.uk/publications/docs/report-on-the-outcomes-from-a-randomised-controlled-trial1.pdf> (accessed 28 February 2015).
- McNeill F (2002) Beyond what works: How and why do people stop offending?.
- Miles MB and Huberman AM (1994) *Qualitative Data Analysis*. Thousand Oaks, CA.: Sage.
- Moore R (2015) A compendium of research and analysis on the offender assessment system (OASys) 2009-2013. Ministry of Justice. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/449357/research-analysis-offender-assessment-system.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf) (accessed 20 April 2020).
- Neyroud P (2015a) Rethinking the gateway: using evidence to reform the criminal justice system for victims and people who offend, pp. 1–28. Howard League for Penal Reform, October 2015. Available at: [https://d19ylpo4aovc7m.cloudfront.net/fileadmin/howard\\_league/user/pdf/Publications/Gateway\\_final\\_web.pdf](https://d19ylpo4aovc7m.cloudfront.net/fileadmin/howard_league/user/pdf/Publications/Gateway_final_web.pdf) (accessed 05 January 2020).
- Neyroud P (2015b) Evidence-based triage in prosecuting arrestees: testing an actuarial system of selective targeting. *International Criminal Justice Review* 25(1): 117–131.

- Neyroud PW (2017) *Learning to Field Test in Policing: Using an analysis of completed randomised controlled trials involving the police to develop a grounded theory on the factors contributing to high levels of treatment integrity in Police Field Experiments*. Doctoral dissertation, University of Cambridge, UK.
- Neyroud P (2018) *Out of court disposals managed by the police: a review of the evidence. Commissioned by the National Police Chief's Council of England and Wales*. University of Cambridge. Available at: <https://www.npcc.police.uk/Publication/NPCC%20Out%20of%20Court%20Disposals%20Evidence%20assessment%20FINAL%20June%202018.pdf> (accessed 27 May 2021).
- Neyroud PW and Slothower MP (2013) *Operation Turning Point: Interim Report on a Randomised Trial in Birmingham*. Cambridge: Institute of Criminology.
- Neyroud P, Slothower M, Sherman L, et al. (2015) *Operation Turning Point (storyboard)*, College of Policing. Available at: [https://whatworks.college.police.uk/Research/Research-Map/Documents/TP\\_Storyboard.pdf](https://whatworks.college.police.uk/Research/Research-Map/Documents/TP_Storyboard.pdf) (accessed 20 February 2021).
- NHS (2015) Alcohol Use Disorders Identification Test (AUDIT) downloaded from. Available at: <https://www.gmw.nhs.uk/download.cfm?doc=docm93jjm4n639.pdf&ver+1017> (accessed 16 October 2015).
- National Police Chiefs' Council (NPCC) (2017) *Charging And Out of Court Disposals: A National Strategy*. NPCC (accessed 20 February 2019).
- Office of National Statistics (ONS) (2017) Crime Severity Score data tool. ONS.gov.uk. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/crimeseverityscoredatatool> (accessed 05 January 2018).
- Oyserman D and Markus HR (1990) Possible selves and delinquency. *Journal of Personality and Social Psychology* 59(1): 112.
- Parker J and Grasmick HG (1979) Linking actual and perceived certainty of punishment: an exploratory study of an untested proposition in deterrence theory. *Criminology* 17(3): 366–379.
- Paternoster R (1987) The deterrent effect of the perceived certainty and severity of punishment: a review of the evidence and issues. *Justice Quarterly* 4(2): 173–217.
- Punch M (1986) *The Politics and Ethics of Fieldwork*. Beverley Hills, CA: Sage.
- Rex S (1999) Desistance from offending: experiences of probation. *Howard Journal of Criminal Justice* 38(4): 366–383.
- Routledge G (2015) *A Protocol and Experimental Trial: The Checkpoint Desistance Programme in Durham*. Cambridge: Institute of Criminology.
- Sadler G (2010) *Evaluation of the Impact of the HM Prison Service Enhanced Thinking Skills Programme on Reoffending. Outcomes of the Surveying Prisoner Crime Reduction (SPCR) Sample*. Ministry of Justice Research Series 19/10. Available at: <https://www.justice.gov.uk/downloads/publications/research-and-analysis/moj-research/eval-enhanced-thinking-skills-prog.pdf> (accessed 27 May 2021).
- Sampson RJ and Laub J (1997) A life-course theory of cumulative disadvantage and the stability of delinquency. In: Thornberry TP (ed.) *Developmental Theories of Crime and Delinquency*. New Brunswick, NJ: Transaction Publishers, pp. 133–163.
- Sherman LW (2011) Offender desistance policing (ODP): less prison and more evidence in rehabilitating offenders. In: Bliesener T, Beelmann A and Stemmler M (eds) *Antisocial Behavior and Crime: Contributions of Developmental and Evaluation research to Prevention and Intervention*. Cambridge, MA: Hogrefe Publishing, pp. 199–218.

- Sherman LW and Neyroud PW (2012) *Offender-Desistance Policing and the Sword of Damocles*. London: Civitas.
- Sherman LW, Neyroud PW and Neyroud E (2016) The Cambridge crime harm index: measuring total harm from crime based on sentencing guidelines. *Policing: A Journal of Policy and Practice* 10(3): 171–183.
- Torgerson DJ and Torgerson CJ (2008) *Designing Randomised Trials in Health, Education and Social Services*. Basingstoke: Palgrave Macmillan.
- Trotter C (1996) The impact of different supervision practices in community corrections: cause for optimism. *Australian & New Zealand Journal of Criminology* 29(1): 1–19.
- Trotter C (2000) Social work education, pro-social orientation and effective probation practice. *Probation Journal* 47: 256–261.
- Urwin S (2016) Cambridge CRIM-PORT. Retrieved from Algorithmic forecasting of offender dangerousness for police custody officers: an assessment of accuracy for the Durham Constabulary Model. Available at: <http://www.crim.cam.ac.uk/alumni/theses/Sheena%20Urwin%20Thesis%2012-12-2016.pdf> (accessed 5 January 2018).
- Weir K, Routledge G and Kilili S (2019) Checkpoint: an innovative programme to navigate people away from the cycle of reoffending: implementation phase evaluation. *Policing: A Journal of Policy and Practice* 15(1): 508–527. Available at: <https://doi.org/10.1093/police/paz015> (accessed March 2021).
- Williams KR and Hawkins R (1986) Perceptual research on general deterrence: a critical review. *Law and Society Review* 20: 545–572.