THE CAMPBELL COLLABORATION

Protocol:
Third Party Policing for Reducing Crime and Disorder: A Systematic Review
Lorraine Mazerolle, Elizabeth Eggins & Angela Higginson

Submitted to the Coordinating Group of:

☒	Crime and Justice
☐	Education
	☐ Disability
☐	International Development
	☐ Nutrition
☐	Social Welfare
☐	Other:

Plans to co-register:

☒	No		
☐	Yes	☐ Cochrane	☐ Other
☐	Maybe		

Date Submitted: 16th April 2014
Date Revision Submitted: 14th June, 2016
Approval Date: 22nd September, 2016
Publication Date:

Note: Campbell Collaboration Systematic Review Protocol Template version date: 24 February 2013

BACKGROUND

The Problem, Condition or Issue

Criminological research universally shows that crime and disorder problems are largely driven by a small percentage of people, clustered in specific places, and committed at particular times of the day, week, month and year. Years of research shows that a substantial proportion of crime is generated by just 3–15% of offenders (Allard, Chrzanowski, & Stewart, 2012; Thornberry, Huizinga, & Loeber, 1995; Wolfgang, 1973), who have very well understood types of antecedents and pathways to offending (see Jennings & Reingle, 2012; Moffit, 1993, 2003; Piquero, 2008; Piquero, Jennings, & Barnes, 2012). Placed-based research reveals that a substantial proportion of crime clusters at 3-10% of city street segments (see Sherman, Gartin, & Buerger, 1989; Weisburd et al., 2004, 2012), a trend that holds for street-level drug crime (Weisburd & Green, 1995; Weisburd & Mazerolle, 2000) and violent crime (Braga, Papachristos, & Hureau, 2010; Eck, Gersh, & Taylor, 2000), with even more concentrated clusters for juvenile crime (Weisburd, Morris, & Groff, 2009). In addition to clustering by place and offender, there is also a predictable temporal pattern to crime, whereby crime and disorder tend to cluster in the evening (Felson & Poulsen, 2003; Townsley, 2008), over the weekend (Uittenbogaard & Ceccato, 2012), during winter for robberies (van Koppen & Jansen, 1999) and during summer for violent crime (Farrell & Pease, 1994). These various forms of crime clustering suggest that crime and disorder problems are somewhat predictable, providing police with a wide range of opportunities to focus their crime control and prevention strategies.

Third Party Policing (TPP) is identified as one of eight key policing innovations of the 21st century (Weisburd & Braga, 2006) that enables police to target the places, people and/or times that disproportionately contribute to crime and disorder problems. TPP expands the capacity of police to target crime and disorder clusters in two distinct ways: (1) by creating a partnership between police and non-police third parties that (2) harnesses the third party's resources and legal powers to control or prevent a crime or disorder problem. In TPP, police partner with external entities ('third parties') – such as government regulators and inspectors, housing authorities, licensing authorities, and business owners – to harness the partner's legal powers and responsibilities to regulate or alter the underlying social, physical, temporal and/or situational conditions that generate crime and disorder problems (Buerger & Mazerolle, 1998; Green-Mazerolle & Roehl, 1998; Mazerolle & Ransley, 2005). In TPP, the police indirectly address crime and disorder problems by working through (and with) their third party partners and those partners' range of legal levers.

The trend towards partnership approaches in policing, such as TPP, emerged from global transformations in governance and regulation during the 1990s (see Mazerolle & Ransley, 2005 for review). These transformations generated a proliferation of regulatory agencies and laws (Braithwaite, 1999, 2000), blurring the boundaries between traditional categories of law (Cheh, 1998; Mazerolle & Ransley, 2005). For policing, these blurred boundaries and

broad regulatory networks created opportunities for partnerships with external crime control 'nodes' or entities (Crawford, 2006, 2009; Crawford et al., 2005; Ericson, 2007; Jones & Newburn, 2006; Loader, 2000; Shearing, 2007). As a result, in many police jurisdictions throughout the world, the presumption is now that police will use partnerships to control crime and disorder problems. In Scotland, for example, Section 32 of the Police and Fire Reform (Scotland) Act 2012 embodies the notion of collaboration and partnership in legislation. Similarly, the UK Crime and Disorder Act 1998 mandates that the police forge partnerships to control crime and disorder problems (see also Independent Police Commission, 2013).

Legal reforms that shift the notion of partnerships in policing from being encouraged to being mandated is one driver of partnerships in policing. Other drivers for partnerships in policing are the proliferation of crime control programs in police departments (e.g., see Weisburd & Eck, 2004), ad-hoc or episodic initiatives developed at the grassroots of policing (e.g., see Scott, 2013), and political directives for partnerships (e.g., in counterterrorism, see Bayley & Weisburd, 2009; Brewer, 2013). TPP is a form of policing that is differentiated from other partnership approaches (e.g., grassroots initiatives, political directives and crime control programs such as Pulling Levers, Plural Policing, Problem-Oriented Policing) which analyse problems and then activate mechanisms of change in a range of ways that may or may not include partnerships with entities who possess legal levers. For example, Pulling Levers is a focused deterrence approach that entails identifying a crime problem; forming an interagency working group; researching the characteristics of the identified problem; devising a response that includes a range of sanctions (e.g., police crackdowns); mobilising community resources and social services to complement police responses; and consistent communication directly with offenders so that they understand the action being taken by police and other agencies (see Braga & Weisburd, 2012). By contrast, the TPP mechanism of change requires a partnership between police and an enitity with one or more legal levers (i.e., a third party) such that the form of the intervention is defined by the explicit activation or escalation of the legal processes delineated by the third party's legal lever(s).

The partnerships, in TPP, are grounded by existing legal processes: they are *formed* because of the legal provisions available to third party partners and *sustained* because the legal provisions offer mutual benefit to both police and their third party partners. TPP is a policing approach that remains part of the "new crime control establishment" (Garland, 2001, p. 17) that is relevant to policing in times of fiscal restraint (e.g., see Ayling, Grabosky, & Shearing, 2009) and consistent with the trend towards proactive focusing of police resources on clusters of criminogenic places, people and situations (e.g., see Lum, Koper, & Telep, 2011; Sherman, 2013; Telep & Weisburd, 2012; Weisburd & Eck, 2004). In a preliminary review of the extant evaluation literature, Mazerolle and Ransley (2005) concluded that TPP may be effective for reducing a wide range of crime and disorder problems. In this current review, we will update, refine and expand Mazerolle and Ransley's (2005) previous work to systematically assess the effectiveness of TPP for reducing crime and disorder problems.

The Intervention

TPP is a policing approach that requires a partnership between police and a third party. The third party is valuable to police because they have access to legal provisions (i.e., legal levers) that are (or could be) applied to control or prevent a crime or disorder problem. Figure 1 illustrates the components of a TPP intervention. As seen Figure 1, TPP involves three key players: (1) the police ('first party'), (2) a crime or disorder problem ('second party'), which could be a problem place, problem people, or a situation where criminogenic places, times and people converge, and (3) an external entity ('third party') that police partner with to control or prevent the crime or disorder problem. We describe each of these component parts below.

In TPP, the 'first party' is defined as the public police. As Figure 1 illustrates, public police work in partnership with a third party for the purposes of controlling or preventing a crime and/or disorder problem. Partnerships may be forged in an ad-hoc episodic manner (see Mazerolle & Ransley, 2005), through a program of crime control activities (e.g., Pulling Levers Policing, see Braga & Weisburd, 2012; Problem-Oriented Policing, see Weisburd et al., 2010), and/or because the partnership is mandated by law (e.g., the UK Crime and Disorder Act, 1998; the Scottish Police and Fire Reform Act of 2012).
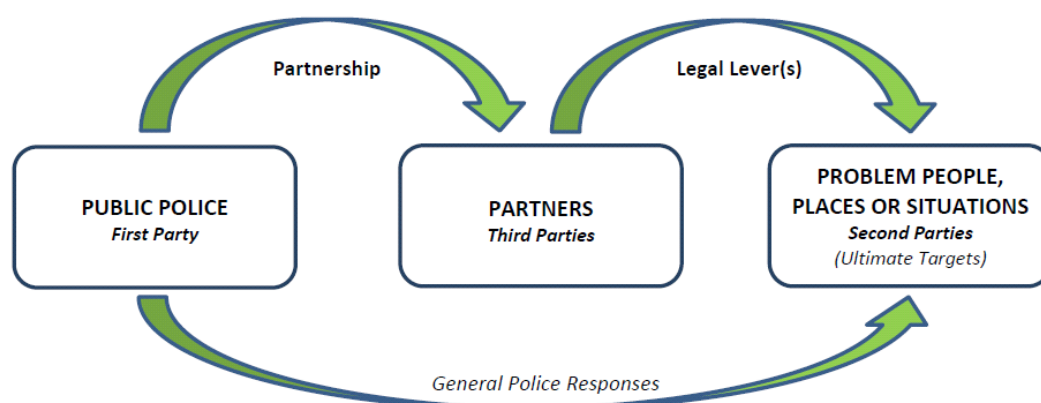

Figure 1. Third Party Policing Model

The 'second party' in TPP is defined as the ultimate crime control or prevention target (see Buerger & Mazerolle, 1998; Mazerolle & Ransley, 2005). Consistent with Routine Activity Theory, the ultimate target of a TPP intervention can be a problem person (a motivated offender), a problem place (an amenable place), or a problem situation (a suitable target, absence of suitable controllers) (see Cohen & Felson, 1979; Eck, 1994; Felson, 1995). In essence, TPP interventions aim to focus police resources on one or more criminogenic

factors that either allow crime problems to flourish, or prevent crime problems from emerging or escalating.

The 'third party' lies at the centre of the TPP intervention approach. A third party is an entity – a person, an agency, organisation, or business – operating within a legal framework and with legal powers and responsibilities not directly available to police. The third party is thus the partner and agent of crime control within TPP. A third party can be an individual (e.g., a bar staff member, property owner), an organisation (e.g., Pharmacy Guild), a business (e.g., a bar), a regulatory authority (e.g., liquor licensing authority, local council, school), a government department (e.g., education department), or a network of collaborating agencies (e.g., see Green, 1996), all of which have statutory responsibilities that are unavailable to police.

How the Intervention Might Work

The key defining feature of TPP is that police indirectly, rather than directly, target crime and disorder problems, and they do so through a partnership with a third party with access to a legal lever. We hypothesise that TPP controls and/or prevents crime through police–third party partnership formation which, in turn, activates, escalates or re-directs the use of existing legal levers to address crime problems Figure 1 (above) illustrates the two critical mechanisms of TPP (represented by the two upper green arrows in Figure 1) as being (a) the character of the crime control partnership between police and the third party, and (b) the process of activating, escalating or re-directing the use of legal levers. These two intrinsically tied mechanisms distinguish TPP from other types of policing and define the underlying mechanisms of TPP interventions.

We hypothesise that it is the *combination* of the partnership and legal lever components that underlies TPP's potential for preventing and controlling crime and/or disorder problems. Specifically, we hypothesise that the TPP mechanism that impacts crime control outcomes is the process of police forming legitimate partnerships with third parties, which then enables the activation (escalation or re-direction) of legal levers. In the subsections that follow, we first describe the partnership and legal lever components. Second, we discuss how the partnership and legal lever components, in combination, are hypothesised to explain the effectiveness of TPP interventions in controlling crime and/or disorder.

The Character of TPP Partnerships

The first necessary mechanism underlying TPP interventions is the dynamic character of the TPP partnership. TPP partnerships exist when police and one or more third parties work together to control or prevent a crime/disorder problem through the initiation and/or escalation of third parties' legal levers. The TPP Partnership Matrix in Figure 2 below captures the different types of TPP partnerships. The matrix categorises TPP partnerships along two continuums: the engagement continuum and the number of third parties within the TPP intervention. The engagement continuum reflects the range of engagement strategies – from collaborative to coercive – that are used to forge and maintain police–third

party partnerships in the effort to control or prevent crime and disorder problems.

The way police engage with third parties will depend on the willingness and capacity of third parties to partner with police to address crime and disorder problems. In some instances, police may need to induce a recalcitrant or less-than-willing third party's cooperation to address a crime and disorder problem. In other instances, third parties may be willing and enthusiastic to cooperate and work with police to address a crime and disorder problem. Mazerolle and Ransley (2005) describe coercive engagement as a 'sledge-hammer' approach that is characterised by forceful engagement techniques. By contrast, the 'carrot,' or collaborative end of the continuum, uses persuasive techniques of engagement. We extend this description and define coercive techniques as tactics or strategies that police use to forge or maintain TPP partnerships where the police either threaten or actually impose negative consequences or remove benefits in order to compel a third party to cooperate (see Raven, 2008). In contrast, we define collaborative techniques as tactics or strategies that are characterised by more consultative or amicable processes that aim to engender willing cooperation from a third party. Scott (2013) provides several examples of the range of partnership engagement techniques that have been implemented by police in past interventions (see also Cherney, 2008; Goldstein & Scott, 2005). For example, on the coercive end of the continuum, police may compel third parties to cooperate and take responsibility for a crime and disorder problem by withdrawing services, publicly shaming the third party, or initiating civil actions against the third party for failure to meet their statutory responsibilities (e.g., bar owners that serve alcohol to minors). Alternatively, police may take a collaborative approach with their partners by making informal requests or appeals for cooperation, educating third parties to increase awareness of their responsibilities, providing incentives or rewards for cooperation (see Grabosky, 1995; Farrell & Roman, 2006), or brokering formal partnerships that are based on cooperative problem-solving, joint decision-making and sharing of resources (see Bond & Gittell, 2010; Claiborne & Lawson, 2005).
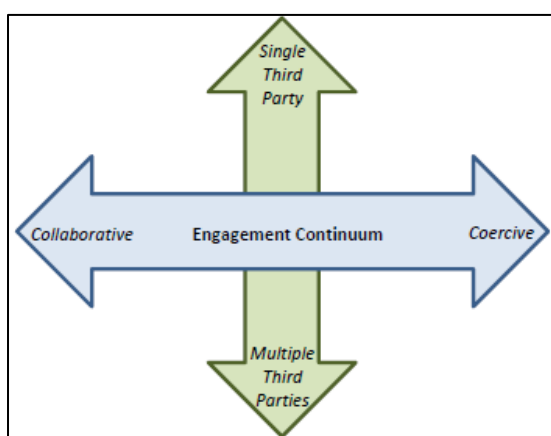
Figure 2. Third Party Policing Partnership Matrix

The first quadrant in Figure 2 is where police and single third parties forge partnerships characterised by collaborative engagement strategies (top, left quadrant). An example of a TPP intervention in this single third party, collaborative quadrant is the DART (Drug Abatement Response Team) intervention, which aimed to address drug-related crime and disorder at residential properties in San Diego (Eck & Wartell, 1998). The police partnered with the City's Code Compliance Department (single third party) who, through the enforcement of nuisance abatement legislation (legal lever), could close properties for up to one year, or fine property owners if persistent drug activity was found at the property. Before resorting to these outcomes, the police and DART advised property owners/managers of the crime problem at their property and the consequences if they did not take steps to alleviate the problem, and met with property owners/managers to devise a plan of action. In other words, police collaboratively partnered with the City Code Compliance Department (third party) to utilise their legal powers, including property closure to remove the 'place' enabling the crime or activating a place manager to deter drug-related crime and disorder, to modify the conditions underlying the crime problem.

The second quadrant is where the police focus on a single third party and use coercive engagement techniques (top, right quadrant) where there is non-compliance. Ransley and colleagues (2011) describe the relationship between police and rogue pharmacists as being a single third party coercive TPP partnership. In Queensland (Australia), police can use a real-time recording database (Project STOP) to identify anomalies in the way pharmacies record (or fail to record) purchase information about products containing pseudoephedrine (e.g., customer identification details). In Queensland, under several statutory and regulatory provisions, it is compulsory for pharmacists (third party) to record information about purchasers of controlled substances and share information with both police and health authorities. Failure to fulfil the mandatory reporting obligations can result in criminal charges and civil penalties (e.g., loss of licence to sell controlled substances). These legal provisions can be used by police to coerce rogue pharmacies to cooperate with police initiatives.

The third quadrant is where the TPP partnership involves multiple third parties and the engagement techniques used to forge and maintain the partnership are collaborative (bottom, left quadrant). An example of a TPP intervention involving multiple third parties

working in a collaborative manner is the Specialised Multi-Agency Response Team (SMART) intervention (Green, 1995, 1996). In this study, addresses with high numbers of calls-for-service or drug-related arrests received a TPP intervention where police encouraged property managers to initiate legal levers by discussing the drug crime and disorder problem at their property, reminding them of the legal levers they were responsible for implementing (e.g., evicting problem tenants under Drug Nuisance Abatement laws and abiding by housing, safety, health and fire legislation), and offering a free course on property management. Police also partnered with SMART, which comprised city inspectors from various regulatory agencies (e.g., Housing, Public Works) responsible for enforcing legal levers in their respective areas. After being provided information by police regarding problem properties, the SMART inspectors escalated, where there was non-compliance, the regulatory legislation by issuing code violations that could result in fines or property closure.

The fourth quadrant is where the TPP partnership involves multiple third parties and the engagement techniques used to forge and maintain the partnership fall at the coercive end of the Engagement Continuum (bottom, right quadrant). In this context, police may need to compel one or more recalcitrant third parties who are unwilling or unable to work with them to prevent or control a crime and disorder problem. For example, in an effort to reduce alcohol-related crime and disorder problems in Wisconsin, police used the media to publicly shame licensees and public officials (third parties) who were lax in their implementation and enforcement of liquor licensing legal levers (cited in Scott, 2013; see also Green Bay Police Department, 1999). As a result, third parties became stricter in their implementation and enforcement of liquor licensing legal levers, which then assisted police with the regulation of conditions underlying alcohol-related crime and disorder (e.g., public intoxication).

Evidence supporting the effectiveness of partnership approaches for addressing crime and disorder is growing (see Berry et al., 2011 for review). Rosenbaum (2002) provides a comprehensive list of reasons that may explain why a partnership approach is particularly effective for addressing complex crime and disorder problems. Among these reasons is the increased capacity for partnerships to target criminogenic risk factors in a multifaceted way while at the same time pooling and executing resources efficiently (see also Cherney, 2008; Rosenbaum & Schuck, 2012). Gittell (2006) provides a relational perspective by suggesting that multiagency partnerships are conducive to the development of 'relational coordination' which is characterised by "frequent high-quality communication supported by relationships of shared goals, shared knowledge, and mutual respect" (Bond & Gittell, 2010, p. 119). Gittell has empirically demonstrated that the level of relational coordination between multiagency partners impacts partnership efficiency and attainment of desired outcomes (Gittell, 2011; Gittell, Fairfield et al., 2000; Gittell, Seidner, & Wimbush,2010; Gittell, Weinberg et al., 2010). The parallel between these explanations for the effectiveness of partnerships and TPP is clear. TPP partnerships provide a forum for pooling resources, for targeting criminogenic risk factors in a multifaceted way through different legal levers, and for generating relational coordination through communication and relationships characterised by shared goals and knowledge. However, as will be explored below, the partnership component is insufficient, in

isolation, for understanding how TPP interventions may work to prevent crime and disorder.

Legal Levers

The second necessary mechanism of TPP interventions is the activation, escalation or re-direction of a third party's legal levers. Legal levers are broadly defined as the legal powers possessed by third parties that create a crime control or crime prevention capacity that is otherwise unavailable to police. Police use TPP partnerships to access, influence, activate, escalate or re-direct these legal levers in TPP interventions. Examples of legal levers include conduct licensing (e.g., alcohol, firearms), mandatory reporting (e.g., chemical sales, child abuse), orders to control behaviour (e.g., gang or domestic violence injunctions, truancy regulations), orders under regulatory codes (e.g., building, fire, health and safety, noise codes), and property controls (e.g., drug nuisance abatement). Legal levers define and shape TPP interventions. First, by specifying third parties available for police partnership. Second, the procedural aspects of a TPP intervention is based on the legally mandated processes, possible legal outcomes, or consequences of the legal levers available to police through third parties.

Legal levers are positioned within an overarching legal framework that aims to regulate social, economic or functional activities in a given jurisdiction (e.g., health and safety, licensing, banking, transport). Third parties have the legal power to regulate these activities within their jurisdiction through the implementation or enforcement of legal levers. For example, a bar owner implements legal levers around responsible service of alcohol in order to regulate the behaviour of patrons (e.g., staff training, alcohol serving times, age restrictions). In contrast, a liquor licensing authority enforces compliance with licensing conditions, also aimed at regulating behaviour of patrons and licensed establishments (e.g., fines for serving underage patrons). In TPP, police assume that conditions that allow a crime problem to flourish can be controlled when (or if) a third party uses their legal lever to regulate behaviour, whether that be individuals, groups of individuals, or characteristics of places or geographic areas. For example, if a school-age person is committing crimes during the day time, police might encourage schools to activate and/or escalate their truancy laws to pressure the young person to attend school. Thus, police partner with third parties to modify the criminogenic conditions underlying a crime problem, thereby indirectly controlling or preventing the problem through implementation, or enforcement, of available legal levers (see Mazerolle, 2014).

Legal levers can be categorised by (a) their source of legal authority, (b) extent of their application, and (c) type of legal outcomes or consequences they may produce. Sources of legal authority include statutes, regulation/subordinate legislation, contract or tort, and the extent of application is either general or targeted (e.g., specific population, area, parties to contract, those with duty of care). Depending on the legal framework and the third party, the types of legal consequences can be criminal, civil, or administrative in nature including fines, licence revocation, incarceration, eviction, property forfeiture, orders for compensation or damages, infringement notices, injunctions, and refusal of entry into or ejection from

licensed premises.

The legal frameworks from which legal levers are drawn dictate the process of TPP interventions[1]. Legal levers are drawn largely from the increasingly complex web of regulatory laws in the 'new regulatory state' (Braithwaite, 2002) where the emphasis is not so much on post-event use of formal legal sanctions, but rather on articulated and graduated actions that ultimately seek voluntary cooperation (see Mazerolle & Ransley, 2005 for a review). In explicating the theory of responsive regulation, Ayres and Braithwaite (1992) and Braithwaite (2006, 2011) describe this system of graduated sanctions as a regulatory pyramid (see Figure 3). The pyramid captures how regulators respond to each successive act of non-compliance by progressing through a hierarchical range of sanctions in a systematic and increasingly punitive way. As Ransley (2014) suggests, the range of legal levers that could be used in TPP interventions is extensive. Our preliminary review of TPP literature (see Mazerolle & Ransley, 2005) indicates that most legal levers utilised in TPP interventions align closely with Braithwaite's (2006, 2011) concept of the regulatory pyramid. That is, legal levers are activated by initiation of more benign consequences to encourage compliance (e.g., education, warning letter) and then sequentially escalate to more punitive consequences to coerce compliance (e.g., infringement notices, to fines, to license revocation), with the ultimate sanctions at the tip of the pyramid. It is this codified and stipulated process for regulating conduct that differentiates TPP from other policing processes[2].

Braithwaite (2006, 2011) suggests that regulation of social, economic or functional activities through the pyramid structure is both efficient and effective, provided regulators are willing and able to consistently initiate and escalate sanctions in response to non-compliance. The idea is that the possible range of sanctions aligns with the array of capabilities and motivations that underlie non-compliance (see Figure 3). For example, if an 'offender' is responsive to persuasive, normative requests to comply with rules, he or she is likely to alter their behaviour. In contrast, for a rational actor who is only responsive to the threat of punishment, then the use of persuasion or education at the bottom of the pyramid may not be effective for obtaining compliance. In the responsive regulatory model, this type of offender would be coaxed into compliance with more deterrent-based sanctions further up the pyramid (e.g., warning letters or civil actions). Provided that citizens believe "in the inexorability of escalation if problems are not fixed" (Braithwaite, 2011, p. 489), most escalations should not proceed far beyond the lower levels of the pyramid. Moreover, Braithwaite (2011) suggests that escalations are unlikely to compromise perceptions of legitimacy pertaining to the law or the regulator if the regulatory process begins with approaches that align with the principles of procedural and restorative justice. As a result,

---

[1] Unlike problem-oriented policing, where the process of intervention is driven by analysis of a problem and then selection of a suitable response(s) based on the specific characteristics of the problem (Goldstein, 1979, 1990; Spelman & Eck, 1987).
[2] In addition, unlike other partnership-type policing approaches (e.g., community-oriented, networked, plural or pulling levers policing), it is a necessary condition for TPP that partners possess a legal lever that is otherwise unavailable to police.

the use of responsive regulation is likely to foster voluntary compliance through perceptions of legitimacy. Therefore, responsive regulation is an effective strategy for regulating a range of factors that may underlie crime problems by fostering voluntary compliance and also responding to non-compliance in a way that addresses the full range of motivations underlying offending.
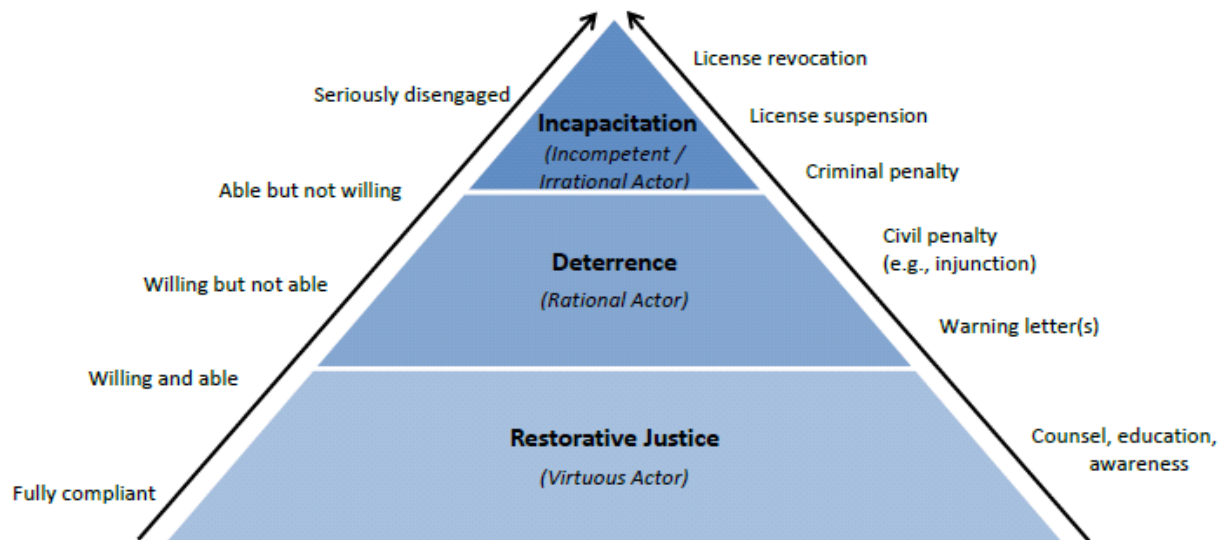

Figure 3. Regulatory Pyramid (adapted from Ayres & Braithwaite, 1992; Braithwaite, 2006, 2011).

The Combined Effect of TPP Partnerships and Legal Levers

In the previous subsections, we described how partnerships and legal levers can help control or prevent crime and disorder problems. We argue, however, that it is the *combination* of these two necessary factors that explains how TPP interventions may work to control or prevent crime and disorder problems. We suggest that the formation of police–third party partnerships fosters the activation (escalation or re-direction) of legal levers, which then enables the control or prevention of crime and disorder problems.

Why is a partnership alone not enough? We suggest that a partnership between police and another entity on its own is less likely to generate the capacity to control crime without the third party having access to a pre-existing legal lever. In TPP, the addition of a legal lever structures, legitimises and prioritises the partnership between police and third parties. A range of research demonstrates how cross-agency partnerships are more likely to be effective when there is a clear structure to the partnership, including the articulation of roles, responsibilities and processes (e.g., see Berry et al., 2011; Foster-Fishman et al., 2001; Meyer & Mazerolle, 2013; Rosenbaum & Schuck, 2012; Roussos & Fawcett, 2000; Zakocs & Edwards, 2006). We propose that legal levers provide a framework for structuring partnerships because they pre-establish roles and responsibilities and procedural aspects of an intervention. Moreover, because legal levers mandate the legal responsibilities of third parties, a partnership with police to control or prevent a crime problem through the activation, escalation or redirection of legal levers is legitimised and more likely to be prioritised by third parties.

Why are legal levers alone not enough? If legal levers are already positioned within regulatory pyramids, they should already be controlling or preventing crime and disorder problems by promoting voluntary compliance in the way hypothesised by responsive regulation theory. However, we suggest that legal levers are not consistently activated (escalated or re-directed) by third parties in a way that makes them effective for controlling or preventing crime and disorder problems. Indeed, legal levers are more often than not created without reference to their potential as a tool for crime prevention or control. Street-level bureaucrat literature highlights how those on the 'front-line' of policy, regulatory and legislative implementation (i.e., third parties) can lack knowledge of legal provisions available to them, and even if they are aware of the provisions, a variety of factors influence how street-level bureaucrats use their discretion to implement legal provisions (Gofen, 2013; Lipsky, 2010; Tummers, 2011). Third parties may know little about the activation procedures delineated by legal levers, inconsistently activate legal levers, or lack the capacity to do so (e.g., see Baldwin & Black, 2008; Weber, 2013). As a result, the patchwork of individual attitudes, levels of knowledge and beliefs amongst third parties influences the way they activate legal levers in day-to-day practice and, in turn, the potential for legal levers to regulate factors that underlie crime and disorder problems. In TPP, we hypothesise that the formation of a partnership between police and third parties impacts the way third parties perceive legal levers and their capacity to consistently activate legal levers.

Figure 4 explicates the hypothesised process by which TPP partnerships and legal levers may work to impact crime and disorder problems. First, the presence of a legal lever prioritises, legitimises and structures the partnership between police and third parties. Second, the formation of TPP partnerships augments the ability of legal levers to reduce crime and disorder by (a) impacting third parties' capacity to consistently and reliably activate their legal levers, and (b) altering third parties' perceptions of their legal levers. That is, the TPP partnership increases the potential for third parties to activate the full range of their legal levers in the way envisioned by Braithwaite's regulatory pyramid (from education and persuasion through to prosecution). Third, the consistent and reliable activation of legal levers increases the likelihood that the wide range of complex motivations underlying compliance, and the criminogenic factors underlying crime problems, are effectively regulated, thereby ultimately impacting levels of crime and disorder.
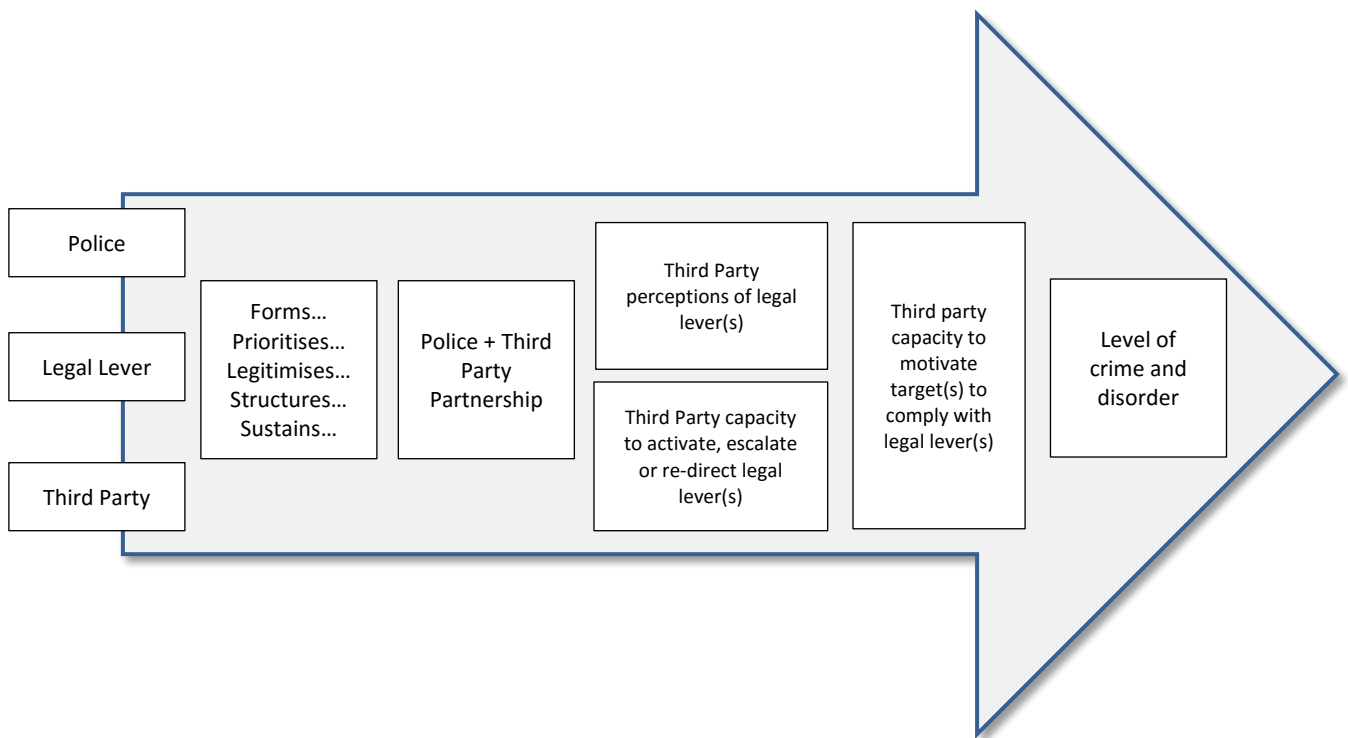

Figure 4. Logic model depicting how TPP impacts crime and disorder.

Why it is Important to do the Review

In a non-Campbell Collaboration review, Mazerolle and Ransley (2005) used systematic review techniques to locate, assess and describe the extant TPP evaluation literature. The authors identified a large pool of studies that varied in terms of methodological rigour, type of third party, type of legal lever and type of crime problem targeted by the intervention. On the basis of mostly positive effect sizes across individual studies, Mazerolle and Ransley

concluded that TPP appeared to be an effective policing strategy for reducing a wide range of crime and disorder problems.

Almost ten years have passed since Mazerolle and Ransley's (2005) review. At the time of their review, TPP terminology was only just beginning to emerge in the literature and was not yet part of the general policing lexicon. Although TPP has since been identified as one of eight key policing innovations of the 21st century, the approach is not without critique (e.g., see Desmond & Valdez, 2013; Meares, 2006). Therefore, we argue that an updated and broader systematic review of TPP's effectiveness for reducing crime and disorder is required. The review we propose will enhance Mazerolle and Ransley's previous work by including a more expansive search of published and unpublished literature and, provided sufficient data is available, will include a meta-analysis to determine the effectiveness of TPP for reducing crime and disorder. As we note above, the TPP approach aligns closely with the trend toward partnerships in policing and the focusing of police tactics on people, places and situations. The results of our review will assist policy makers and practitioners to make informed decisions about how TPP can be used to focus their resources, use their existing legal levers, and build partnerships to address crime and disorder.

OBJECTIVES

The primary objective of this review is to systematically evaluate the impact of TPP interventions on crime and/or disorder. We will achieve this by synthesising the results of published and unpublished empirical research on TPP interventions and by addressing the following research questions:

1. What impact does TPP have on crime and/or disorder?

2. Does the impact of TPP vary by the type of TPP partnership?

3. Does the impact of TPP vary by the type of legal lever or third party utilised?

4. Does the impact of TPP vary by the type of crime or disorder targeted?

5. Does the impact of TPP vary by the target of the TPP intervention (e.g., offenders versus crime places)?

METHODOLOGY

CRITERIA FOR INCLUDING STUDIES IN THE REVIEW

Types of study designs

Our review will consider quantitative research that uses randomised experimental (e.g., RCTs) or 'strong' quasi-experimental evaluation designs with a valid comparison group that does not receive the intervention. In most instances, we expect that the control group or

comparison condition will be 'business-as-usual'. For example, police districts operating in their usual fashion are compared to experimental police districts that implemented a specific TPP intervention. However, we will also accept designs where the comparison group receives no intervention or an alternative intervention (treatment-treatment designs).

Although not as robust as randomised experimental designs, strong quasi-experiments can be used to provide causal inference when the nature of the design attempts to minimise threats to internal validity. This can be achieved in a number of ways, such as: controlling the assignment of cases to treatment and control groups (regression discontinuity), matching the characteristics of the treatment and control groups (matched control), statistically accounting for differences between the treatment and control groups (designs using multiple regression analysis), or providing a difference-in-difference analysis (parallel cohorts with pre-test and post-test measures).

We will include the following quasi-experimental designs in our synthesis of the effectiveness of TPP:

- Regression discontinuity designs
- Designs using multivariate controls (e.g., multivariate models that control for confounding factors whilst also examining the effects of group membership)
- Matched control group designs with or without pre-intervention baseline measures (propensity or statistically matched)
- Unmatched control group designs with pre-intervention measures (difference-in-difference analysis)
- Short interrupted time-series designs with control group (less than 25 pre- and 25 post-intervention observations (Glass, 1997))[3]
- Long interrupted time-series designs with or without a control group (≥25 pre- and post-intervention observations (Glass, 1997))

To address potential bias due to research design, we will perform a subgroup analysis using research design as a predictor variable. In addition, time-series designs will be synthesised separately because the effect size has a different meaning than numerically equivalent effects sizes for other quasi-experimental designs.

---

[3] In distinguishing between and pre- and post-test designs with control groups and short interrupted time-series designs with control groups, the key factor is whether the study reports on data from a group of subjects (e.g., offenders) or a single subject (e.g., police district, region). In pre- and post-test control group designs, the outcome is typically reported as a mean value for each of two groups of subjects (treatment and control), calculated at two time points (before and after the intervention). For example, in a pre- and post-test control group design, the study might compare the mean time to reoffending by offenders in the treatment group and compare that to the control group. On the other hand, a short interrupted time-series design with a control group typically reports on data from two subjects, where each subject is a group or area. Each subject is observed repeatedly over time, and one subject receives an intervention during the period of observation. In these studies the outcome is reported as a single measure, rather than as a mean. For example, a study may measure total offences every year over a ten year period for two similar police districts, where one district begins to focus on the use of TPP interventions during the period of observation (e.g., during year 4), and the control area where the intervention never occurs.

To be included in the meta-analysis, evaluations must have also reported an effect-size, or sufficient detail to allow an effect size to be calculated. Where there is not sufficient detail to calculate the effect size and standard error in an otherwise eligible study, we will attempt to contact the corresponding author for the required information.

Types of interventions

To be eligible for inclusion in the review, each piece of research must satisfy the criteria that define a TPP intervention (see Background section above):

- The presence of at least one third party with a legal lever; AND
- The presence of a police partnership with a third party that addresses a crime problem through the use of legal lever(s) accessible to the third party.

A third party is defined as an entity external to the police and can be an individual (e.g., a bar staff member, property owner), an organisation (e.g., Pharmacy Guild), a business (e.g., a bar), a regulatory authority (e.g., liquor licensing authority, local council, school), a government department (e.g., education department), or a network of collaborating agencies (e.g., see Green, 1996). To qualify as a third party for this review, the entity **must** possess a legal lever. A legal lever is defined as the legal power(s) that an entity is responsible for implementing or enforcing in order to regulate social, economic or functional activities in a given jurisdiction (e.g., health and safety, licensing, transport, education). Examples of legal levers include conduct licensing (e.g., alcohol, firearms), mandatory reporting (e.g., chemical sales, child abuse), orders to control behaviour (e.g., gang or domestic violence injunctions, truancy regulations), orders under regulatory codes (e.g., building, fire, health and safety, noise codes), and property controls (e.g., drug nuisance abatement). An eligible TPP intervention will contain a partnership between police and one or more third parties whereby legal levers are used to control or prevent crime problem. There will be no exclusions on the type of third party or legal levers considered for the review, however, an external entity will only be classified as a third party if they possess a legal lever.

Types of outcome measures

Crime and disorder is the primary outcome of interest for this review. To be included in the review, each TPP evaluation must report at least one crime and/or disorder outcome. Due to variation in the way outcomes are measured across the literature (e.g., see Addington, 2009), the scope of outcomes considered for the review will be relatively wide. We plan to conduct meta-analyses separately for conceptually different outcomes (e.g. we will separately analyse the effects of the intervention on violent crime and property crime) and will conduct moderator analyses to determine if different measurement methods (e.g. survey data vs official data vs observations) result in a different estimate of the effect. We will code and analyse all types of crime (e.g., property, violent, drug offences) and/or disorder that use the following measurement methods:

- Official measures of crime (e.g., arrest data, crime rates, calls-for-service data)
- Unofficial measures of crime (e.g., citizen reported crime via interview or survey)
- All types of crime and/or disorder displacement (see Guerette & Bowers, 2009)
- Diffusion of crime control benefits
- Systematic observations of social disorder (e.g., public intoxication, loitering, solicitation, excess noise, drug dealing)
- Systematic observations of physical disorder (e.g., dilapidated or abandoned properties, rubbish, graffiti)
- Citizen- or practitioner-reported observations of social or physical disorder

Types of participants

As we are interested in the impact of TPP on crime and/or disorder in general, we will include, code and analyse research with any type of participant or unit of analysis. For example, individual or place levels of analysis will be eligible for inclusion. However, we will synthesise studies with different levels of analysis separately.

Settings, timeframes and language

We will consider interventions executed in any country or region and will apply no restrictions on language. Our search will be conducted using the English language; however we will not exclude research written in a non-English language. Because the emergence of TPP is intrinsically linked with the transformation of governance towards the end of the twentieth century (see Mazerolle & Ransley, 2005) our review will focus on TPP interventions conducted from 1980 onwards.

Exclusion Criteria

Qualitative research designs and any study that does not fit the inclusion criteria outlined above will be excluded from the review.

SEARCH STRATEGY FOR IDENTIFYING STUDIES

The corpus of literature for this review will be drawn from a large-scale policing intervention database – Global Policing Database (GPD, www.gpd.uq.edu.au). The GPD has been created outside of this review and is a collaboration between Australian researchers at The University of Queensland, Queensland University of Technology, the London Mayor's Office for Policing and Crime (MOPAC), and the College of Policing in the United Kingdom. The database is designed to capture all published and unpublished experimental and quasi-experimental evaluations of policing interventions since 1950 without any restrictions on outcome measures, language of the research, or type of policing intervention.

The GPD is being compiled using systematic search and screening techniques, including an extensive systematic search of published and unpublished literature sources. All unique

records are screened for relevance to policing based on the title and abstract and, if relevant, proceed to a staged full-text eligibility screening process to verify the presence of a quantitative impact evaluation of a policing intervention. The full protocol can be found on the GPD website, however, Appendix A summarises the GPD compilation process and the point at which TPP studies will be extracted, and Appendix B provides the GPD systematic search and screening methodology. We will use the GPD as the primary search location for the TPP search, as by definition the literature evaluating TPP interventions is a subset of the GPD corpus. This approach will also improve the timeliness and cost-effectiveness of the systematic review.

We will also hand search the most recent issues of specific journals not yet indexed (see Table 1). In addition, we will conduct cited reference searches using all eligible studies and seminal TPP publications (see Table 1).

Table 1. Sources for hand searches and cited reference searches

Sources for Hand Searches	Additional Sources for Cited Reference Searches
British Journal of Criminology Crime & Delinquency Criminology Criminology & Public Policy Journal of Criminal Justice Journal of Criminal Law and Criminology Journal of Experimental Criminology Journal of Quantitative Criminology Journal of Research in Crime & Delinquency Justice Quarterly Police Practice & Research Police Quarterly Policing Policing & Society	*Policing places with drug problems* (Green, 1995, 1996) *Third party policing: Theoretical analysis of an emerging trend* (Buerger & Mazerolle, 1998) *Civil remedies and crime prevention* (Green-Mazerolle & Roehl, 1998) *Third party policing* (Mazerolle & Ransley, 2005) *The case for third party policing* (Mazerolle & Ransley, 2006)

DESCRIPTION OF METHODS USED IN PRIMARY RESEARCH

At this stage of the review, we have not identified the full corpus of research that will be eligible for inclusion. However, our preliminary examination of the TPP evaluation literature found a mixture of experimental and quasi-experimental research methodologies (e.g., Baker & Wolfer, 2003; Higgins & Coldren, 2000; Mazerolle, Kadleck, & Roehl, 1998; Mazerolle, Price, & Roehl, 2000; Weisburd & Green, 1995). For example, Eck and Wartell (1998) evaluated the DART intervention using a randomised controlled trial design where residential properties were randomly assigned to a control group ($n = 42$) or to one of two treatment groups ($n = 42$ and $n = 37$). The main outcome measure used was 30 months of post-intervention official crime data which was aggregated into five six-month intervals for each site.

Other authors have evaluated TPP strategies using quasi-experimental techniques. For instance, Green (1995, 1996) used a pre–post unmatched control group design to examine the impact of the Specialised Multi-Agency Response Team (SMART) intervention on a number of crime and disorder outcomes in Oakland, California. Green used a number of outcome measures, including calls-for-service, narcotics arrests, field contact data and systematic observation of physical disorder, and compared outcomes for the intervention sites ($n = 321$) with the overall city before, during and after the SMART intervention.

DETAILS OF CODING CATEGORIES AND DATA EXTRACTION

Full-text eligibility screening

As noted in previous sections, the corpus of research for this review will be drawn from a large-scale policing intervention database called The Global Policing Database (GPD). The GPD will contain documents that report on experimental and quasi-experimental impact evaluations of policing interventions, with no limits on the type of outcome measures (see Appendix A for GPD screening process). This means that for the purposes of the TPP review, we will only need to screen documents for eligibility on the TPP intervention and outcome measure inclusion criteria (see Appendix C for eligibility screening companion), as all documents in the GPD will have already met the study design criteria.

The full-text of all eligible documents in the GPD will be imported into our Microsoft Access systematic review database – *SysReview*. The SysReview database is customisable to individual review requirements and creates a unique record for each document (see Appendix D for screen shots). Using the 'Eligibility Screening' form (see Figure D.2 in Appendix D), the full-text of each document will be screened, using the screening companion in Appendix C, to identify studies that satisfy the following criteria:

- Research conducted from 1980 onwards; AND
- The presence of at least one third party; AND
- The presence of at least one legal lever; AND
- The presence of a police partnership with a third party with the intention of addressing a crime problem through the initiation or escalation of legal lever(s) accessible to the third party
- Uses a crime and/or disorder outcome measure

Documents that are not excluded during this phase will progress to the in-depth coding phase (see below). To ensure consistency in screening decisions, each document coder will screen 30 documents for eligibility and inter-coder agreement will be calculated (percentage agreement between coders that document is eligible). We will accept an inter-coder agreement of ≥95 per cent. If there is less than 95 per cent agreement, we will implement further training and rescreen the group of documents where agreement fell below the 95 per cent threshold. Disagreements regarding the eligibility of training and non-training

documents will be resolved by a discussion between the coders and the review manager.

After the eligibility screening phase has been completed, a list of eligible documents and the inclusion criteria will be distributed to the TPP Advisory Group for perusal to ensure that eligible studies have not been omitted from the review. Although we will have previously contacted policing experts and other study authors for the GPD, the TPP Advisory Group is comprised of practitioners and scholars who may not have been contacted for the GPD (see Appendix E). Any additional studies provided by the Advisory Group will be assessed for eligibility in the same manner as studies retrieved from the systematic search.

Full-text coding and data extraction

A team of trained research assistants will code the documents using the 'TPP Review Full-Text Coding Companion' which details all coding fields (see Appendix F). We will assess coders' understanding of the coding structure and consistency of coding decisions by implementing the same quality control process used for the eligibility screening phase (see above). Given the anticipated large size of this review, complete double-coding will not be feasible. However, a random 10 per cent sample of each coders' work will be double-coded to verify coding reliability and check for coder drift. In addition, we will use two independent coders to conduct duplicate data extraction for effect size coding fields.

Documents will be read in detail and coded according to fields recommended by research synthesists (e.g., Littel et al., 2008). Specifically, data will be extracted for (a) general characteristics of the study (e.g., intervention location); (b) research methodology (e.g., type of comparison group); (c) study quality (see section below); (d) outcome characteristics (e.g., data source); and (e) effect size data. In addition, a range of data will be extracted on the characteristics of the TPP interventions which will serve a dual purpose of informing qualitative descriptions of included studies and proposed subgroup analyses.

Each document may (a) report multiple outcomes for the one intervention or (b) contain multiple studies with multiple outcomes. SysReview allows for this nested data situation by enabling coders to add multiple outcomes for each unique study, and manually add multiple studies within the one document record (see Appendix D, Figure D.3). The results of the eligibility screening and coding phases will be presented in the final review in the form of a PRISMA flowchart (Moher et al., 2009).

 If there is missing data for key coding fields (e.g., intervention components, data required for effect size calculation), we will attempt to correspond with the document's author(s) to obtain the required information.

Criteria for determination of independent findings

We anticipate two issues relating to the determination of independent findings that will need to be addressed in this review. First, documents may report on multiple studies and/or multiple outcomes. Our protocol for this situation will be to allow documents to contribute

multiple effect sizes, but only contribute one effect size for each outcome. If a document provides multiple effect sizes for one outcome, the mean effect size for that outcome will be calculated using Comprehensive Meta-Analysis 2.0 (Borenstein et al., 2005). The second issue of independence is where multiple documents report data from the same evaluation. We will treat dependent studies as a single study and use all sources to calculate effect sizes for each outcome.

Assessment of methodological quality and risk of bias

We will use a modified version of the Campbell Collaboration International Development Coordinating Group (IDCG) Risk of Bias tool to assess the quality of each eligible study (see Appendix G). Rather than allocate a score or index, we will make a qualitative decision regarding the risk of bias for each eligible study. We have chosen this approach because extreme failure in one area of study quality may be more serious than minor breaches across several areas of study quality. We will present the results of study quality assessment using a 'traffic light' format (see de Vibe et al., 2012). We will not exclude studies on the basis of methodological quality or risk of bias; however, we will conduct sensitivity analysis to determine the impact of study quality on the overall findings.

STATISTICAL PROCEDURES AND CONVENTIONS

Methods of synthesis

We will synthesise the effect sizes for each outcome using a random-effects meta-analysis with inverse variance weighting to account for likely heterogeneity in interventions. We will conduct all analyses using Comprehensive Meta-Analysis software (Borenstein et al., 2005). If a study reports multiple effect sizes for the one outcome, we will use the mean effect size for that outcome. We will synthesise the results of time-series studies separately from other experimental and quasi-experimental designs, as time-series designs standardise for variability over time rather than variability over units, resulting in a different scaling (D. Wilson, personal communication, September 20, 2013).

We will only combine results of evaluations if the outcomes are conceptually equivalent. For example, if studies report violent crime and property crime as separate outcomes, we will conduct two separate meta-analyses – one for violent crime outcomes and one for property crime outcomes – as we do not consider that these two outcomes are conceptually equivalent. We will conduct separate meta-analyses for outcomes measured at different levels of analysis (e.g., individual, police district, country). We will present the results of the meta-analysis in forest plots, including 95 per cent confidence intervals for individual studies and the overall effect.

Measures of treatment effect

We will calculate standardised effect sizes and their standard errors in SysReview for the most commonly reported data, as the database has inbuilt calculations with formulae drawn

from Lipsey and Wilson (2001). For less commonly reported data we will calculate standardised effect sizes and their standard errors using the web-based effect size calculator "Practical Meta-Analysis Effect Size Calculator"[4].

For continuous outcomes we will use Hedges' *g* as the measure of effect size, as it includes an adjustment for estimator bias in smaller samples (Borenstein, 2009). If binary outcomes are found we will calculate a log odds ratio as the measure of effect size. Should an outcome be measured across different studies using binary data in some studies and continuous data in others, we will convert all effect sizes and their variances for that outcome to a common metric once the data are entered into Comprehensive Meta-Analysis software (Borenstein et al., 2005). For example, log odds ratios may be converted to Hedges' *g*, and the meta-analysis conducted on all outcomes using Hedges' *g* as the effect size of choice. Following Borenstein and colleagues (2009), we argue that this approach, whilst imperfect, is preferable to conducting two separate meta-analyses with different effect size measures.

Some studies may use an interrupted time-series design with observations at multiple time points before and after the implementation of an intervention in an area and some may use comparison groups in addition to multiple time points. For studies that collect data at multiple time points, we will assume an underlying uniform distribution for violent crime, and a step function for the effect of the intervention on the outcome. We will therefore calculate an average effect size for the time points before the intervention, and an average effect size for the time points after the intervention, and compare the two. We recognise that there are many other ways to deal with this type of time-series data; however, given the research questions and the likely nature of the intervention effect, we believe that this method is the most defensible and parsimonious.

Unit of analysis

The standardised coding sheet contains fields to code both the unit of treatment and the unit of analysis. We will also assess each study for unit of analysis error, as part of the IDCG risk of bias tool. If a study is assessed as suffering from unit of analysis error, we will correct for the standard error and confidence intervals of the studies, using the formula $\text{SE}_{\text{corrected}} = \text{SE}_{\text{uncorrected}} * \sqrt{(1 + (m - 1) * \text{ICC})}$, where m is the number of units in each cluster, if the intra-class correlation (ICC) can be obtained or estimated. We will attempt to contact the author of the study for an accurate measure of the ICC in the first instance. If this is not available, we will estimate the ICC from similar studies, and report clearly how these estimates were derived. We will also conduct sensitivity analyses for the estimates of the ICC.

Missing data

We will use reported statistics such as *t, F, p,* or *z*-values to convert to effect sizes if effect size data are not reported. If data required to compute effect sizes are missing, we will attempt to

---

[4] http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php

contact the authors of the studies to obtain the data required.

Assessment and investigation of heterogeneity

We will test for heterogeneity using I^2, and Q statistics, following Borenstein et al. (2009), and will calculate and report the between studies variance (τ^2). We will code a range of study-level moderators that we expect would have an impact on the effect size. If there is sufficient information available, we will test the effect of key variables on the heterogeneity of the intervention impact, using moderator analysis for categorical predictors and meta-regression for continuous predictors. We will use a random effects model with inverse variance weighting for all moderator analyses. As indicated by the review objectives, we plan to perform moderator analysis on the following variables: type of crime and/or disorder targeted by the TPP intervention (e.g., violent versus property crime); type of TPP partnership; whether the intervention was exclusively TPP or a selected response as part of another type of intervention (e.g. Problem-Oriented Policing); type of legal lever utilised in the TPP intervention; type of third party police have partnered with; and the type of TPP target (e.g., offenders versus crime places). We will distinguish in the final review between a priori planned analyses (those listed in the protocol) and post hoc analyses identified only during the analytic stage.

Sensitivity analysis

We will conduct subgroup analyses in order to assess the impact of study quality and study design. We will use a random effects model with inverse variance weighting for all sensitivity analyses. Using moderator analysis for categorical variables, and meta-regression for continuous variables, we will perform sensitivity analysis on the effect of study quality, publication status, publication year, and geographic level of analysis. We will distinguish in the final review between a priori planned analyses (those listed in the protocol) and post hoc analyses identified only during the analytic stage.

Assessment of publication bias

We will test and adjust for publication bias using a range of approaches suggested in Rothstein, Sutton, and Borenstein (2005); depending on the data collected, this may include funnel plots and trim-and-fill analysis.

Treatment of qualitative research

We will not use qualitative research to evaluate the impact of TPP interventions on crime and/or disorder.

REFERENCES

Addington, L. (2009). Measuring crime. *Oxford Bibliographies Online: Criminology*. doi: 10.1093/obo/9780195396607-0057

Allard, T., Chrzanowski, A., & Stewart, A., (2012). Targeting crime prevention to reduce offending: Identifying communities that generate chronic and costly offenders. *Trends and Issues in Crime and Criminal Justice No. 445*. Retrieved from http://www.aic.gov.au/publications/current%20series/tandi/441-460/tandi445.html

Ayling, J., Grabosky, P., & Shearing, C. (2009). *Lengthening the arm of the law: Enhancing police resources in the twenty-first century.* New York: Cambridge University Press.

Ayres, I., & Braithwaite, J. (1992). *Responsive regulation: Transcending the deregulation debate.* Oxford: Oxford University Press.

Baker, T., & Wolfer, L. (2003). The crime triangle: Alcohol, drug use, and vandalism. *Police Practice and Research, 4*, 47–61. doi: 10.1080/1561426032000059187

Baldwin, R., & Black, J. (2008). Really responsive regulation. *The Modern Law Review, 71*, 59–94. doi: 10.1111/j.1468-2230.2008.00681.x

Bartholomew, B., Gibbs, J., Mazeika, D., Ahlin, E., Joseph, P., & Miller, N. (2009). Trends in police research: A cross-sectional analysis of the 2006 literature. *Police Practice and Research: An International Journal, 10*, 383–407. doi: 10.1080/15614260903255038

Bayley, D. H., & Weisburd, D. (2009). Cops and spooks: The role of police in counterterrorism. In D. Weisburd, T. Feucht, I. Hakimi, L. Mock, S. Perry (Eds.), *To protect and to serve: Policing in an age of terrorism* (pp. 81–99). doi: 0.1007/978-0-387-73685-3_4

Beckman, K., Gibbs, J., Beatty, P., & Canigiani, M. (2005). Miggans, K., & Hart, M. (2006). Trends in police research: A cross-sectional analysis of the 2002 literature. *Police Practice and Research: An International Journal, 6*, 295–320. doi: 10.1080/15614260500206319

Beckman, K., Lum, C., Wycoff, L., & Larsen-Vanderwall, K. (2003). Trends in police research: A cross-sectional analysis of the 2000 literature. *Police Practice and Research: An International Journal, 4*, 79–96. doi: 10.1080/1561426032000059204

Berry, G., Briggs, P., Erol, R., & van Staden, L. (2011). *The effectiveness of partnership working in a crime and disorder context: A rapid evidence assessment* (Research Report No. 52). London: Home Office.

Bond, B. J., & Gittell, J. H. (2010). Cross-agency coordination of offender re-entry, testing collaboration outcomes. *Journal of Criminal Justice*, *38*, 118–129. doi: 10.1016/j.jcrimjus.2010.02.003

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis* (2nd ed., pp. 221–235). New York: Russell Sage Foundation.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis Version 2* (Computer software). Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester, UK: John Wiley & Sons.

Braga, A. A., Papachristos, A., & Hureau, D. (2010). The concentration and stability of gun violence at micro places in Boston, 1980–2008. *Journal of Quantitative Criminology, 26*, 33-53. doi:10.1007/s10940-009-9082-x

Braga, A. A., & Weisburd, D. (2007). *Police innovation and crime prevention: Lessons learned from police research over the past 20 years.* Washington, DC: National Institute of Justice, US Department of Justice.

Braga, A. A., & Weisburd, D. (2012). The effects of 'pulling levers' focused deterrence strategies on crime. *Campbell Library of Systematic Reviews, 8*(3). doi: 10.4073/csr.2012.6

Braga, A. A., Welsh, B .C., Papachristos, A. V., Schnell, C., & Grossman, L. (2014). The growth of randomized experiments in policing: The vital few and the salience of mentoring. *Journal of Experimental Criminology, 10*, 1–28. doi: 10.1007/s11292-013-9183-2

Braithwaite, J. (1999). Accountability and governance under the new regulatory state. *Australian Journal of Public Administration, 58*, 90–97. doi: 10.1111/1467-8500.00077

Braithwaite, J. (2000). The new regulatory state and the transformation of criminology. *British Journal of Criminology, 40*, 222–238. doi: 10.1093/bjc/40.2.222

Braithwaite, J. (2006). Responsive regulation and developing economies. *World Development, 34*, 884–898. doi: 10.1016/j.worlddev.2005.04.021

Braithwaite, J. (2011). Fasken lecture: The essence of responsive regulation. *UBC Law Review, 44*, 475–520.

Brewer, R. (2013). Enhancing crime control partnerships across government: Examining the

role of trust and social capital on American and Australian waterfronts. *Police Quarterly, 16*, 371–395. doi: 10.1177/1098611113488115

Buerger, M. E., & Mazerolle, L. (1998). Third party policing: Theoretical analysis of an emerging trend. *Justice Quarterly, 15*, 301–328. doi: 10.1080/07418829800093761

Cheh, M. (1998). Civil remedies to control crime: Legal issues and constitutional challenges. In L. Green Mazerolle & J. Roehl (Eds.), *Crime Prevention Studies: Vol. 9. Civil remedies and crime prevention* (pp. 45–66). Monsey, NY: Criminal Justice Press.

Cherney, A. (2008). Harnessing the crime control capacities of third parties. *Policing: An International Journal of Police Strategies & Management, 31*, 631–647. doi: 10.1108/13639510810910607

Claiborne, N., & Lawson, H. A. (2005). An intervention framework for collaboration. *Families in Society: The Journal of Contemporary Social Services, 86*, 1–11. doi: 10.1606/1044-3894.1881

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activities approach. *American Sociological Review, 44*, 588–608.

Cooper, H. (2009). Hypotheses and problems in research synthesis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 19–35). New York: Russell Sage Foundation.

Crawford, A. (2006). Networked governance and the post-regulatory state? Steering, rowing and anchoring the provision of policing and security. *Theoretical Criminology, 10*, 449–479. doi: 10.1177/1362480606068874

Crawford, A. (2009). Governing through anti-social behavior: Regulatory challenges to criminal justice. *British Journal of Criminology, 49*, 810–831. doi: 10.1093/bjc/azp041

Crawford, A., Lister, S., Blackburn, S., & Burnett, J. (2005). *Plural policing: The mixed economy of visible patrols in England and Wales.* Bristol, UK: Policy Press.

de Vibe, M., Bjorndal, A., Tipton, E., Hammerstrøm, K., & Kowalski, K. (2012). Mindfulness based stress reduction (MBSR) for improving health, quality of life, and social functioning in adults. *Campbell Library of Systematic Reviews, 8*(3). doi: 10.4073/csr.2012.3

Desmond, M., & Valdez, N. (2013). Unpolicing the urban poor: Consequences of Third-Party Policing for inner-city women. *American Sociological Review, 78*, 117–141. doi: 10.1177/0003122412470829

Eck, J. E. (1994). *Drug Markets and drug places: A case-control study of the spatial structure of illicit drug dealing* (PhD. dissertation). Available from ProQuest Dissertations and Theses database (UMI No. 9514517).

Eck, J., Gersh, J., & Taylor, C. (2000). Finding crime hot spots through repeat address mapping. In V. Goldsmith, P. McGuire, J. Mollenkopf & T. Ross (Eds.), *Analysing crime patterns: Frontiers of practice* (pp. 49–65). Thousand Oaks, CA: Sage Publications.

Eck. J. E., & Wartell, J. (1998). Improving the management of rental properties with drug problems: A randomized experiment. In L. Green Mazerolle and J. Roehl (Eds.), *Crime Prevention Studies: Vol. 9. Civil remedies and crime prevention* (pp. 161–186). Monsey, NY: Criminal Justice Press.

Ericson, R. V. (2007). Rules in policing. *Theoretical Criminology, 11*, 367–401. doi: 10.1177/1362480607079583

Farrell, G., & Pease, K. (1994). Crime seasonality: Domestic disputes and residential burglary in Merseyside 1988–90. *British Journal of Criminology, 34*, 487–498.

Farrell, G., & Roman, J. (2006). Crime as pollution: Proposal for market-based incentives to reduce crime externalities. In K. Moss & M. Stephens (Eds.), *Crime reduction and the law* (pp. 135–55). London: Routledge.

Farrington, D. P. (1983). Randomized experiments on crime and justice. *Crime and justice*, 257–308. doi: 10.1086/449091

Farrington, D. P., & Welsh, B. C. (2005). Randomised experiments in criminology: What have we learned in the last two decades? *Journal of Experimental Criminology*, *1*, 9–38. doi: 10.1007/s11292-004-6460-0

Farrington, D. P., & Welsh, B. C. (2006). A half century of randomised experiments on crime and justice. *Crime and Justice, 34*, 55–132. doi: 10.1086/500057

Felson, M. (1995). Those who discourage crime. In J. E. Eck & D. Weisburd (Eds.), *Crime Prevention Studies Vol. 4. Crime and place* (pp. 53–66). Monsey, NY: Criminal Justice Press.

Felson, M., & Poulsen, E. (2003). Simple indicators of crime by time of day. *International Journal of Forecasting, 19*, 595–601. doi: 10.1016 /S0169-2070(03)00093-1

Foster-Fishman, P. G., Berkowitz, S. L., Lounsbury, D. W., Jacobson, S., & Allen, N. A. (2001). Building collaborative capacity in community coalitions: A review and integrative framework. *American Journal of Community Psychology, 29*, 241–261. doi: 10.1023/A:1010378613583

Garland, D. (2001). *The culture of control: Crime and social order in contemporary society.* Chicago: University of Chicago Press.

Gibbs, J. C., Beckman, K., Miggans, K., & Hart, M. (2006). Trends in police research: A cross-sectional analysis of the 2003 literature. *Police Practice and Research: An International Journal, 7,* 337–361. doi: 10.1080/15614260600919696

Gittell, J. H. (2001). Supervisory span, relational coordination and flight departure performance: A reassessment of post-bureaucracy theory. *Organization Science, 12,* 467–482. doi: 10.1287/orsc.12.4.468.10636

Gittell, J. H. (2006). Relational coordination: Coordinating work through relationships of shared goals, shared knowledge and mutual respect. In O. Kyriakidou & M. Ozbilgin (Eds.), *Relational Perspectives in Organizational Studies: A Research Companion* (pp. 74–94). Cheltenham, UK: Edward Elgar Publishers.

Gittell, J. H., Fairfield, K., Bierbaum, B., Jackson, R., Kelly, M., Laskin, R., ... Zuckerman, J. (2000). Impact of relational coordination on quality of care, postoperative pain and functioning, and length of stay: A nine-hospital study of surgical patients. *Medical Care, 38,* 807–819. doi: 10.1097/00005650-200008000-00005

Gittell, J. H., Seidner, R. B., & Wimbush, J. (2010). A relational model of how high performance work systems work. *Organization Science, 21,* 490–506. doi: 0.1287/orsc.1090.0446

Gittell, J. H., Weinberg, D. B., Pfefferle, S., & Bishop, C. (2008). Impact of relational coordination on job satisfaction and quality outcomes: A study of nursing homes. *Human Resource Management Journal, 18,* 154–170. doi: 10.1111/j.1748-8583.2007.00063.x

Glass, G. V. (1997). Interrupted time series quasi-experiments. In R. M. Jaeger (Ed.), *Complementary methods for research in education* (2nd ed., pp. 589–608). Washington DC: American Educational Research Association.

Gofen, A. (2013). Mind the gap: Dimensions and influence of street-level divergence. *Journal of Public Administration and Theory.* Advance online publication. doi: 10.1093/jopart/mut037

Goldstein, H. (1979). Improving policing: A problem-oriented approach. *Crime and Delinquency, 25,* 236–243. doi: 10.1177/001112877902500207

Goldstein, H. *(*1990*). Problem-oriented policing.* New York: McGraw-Hill*.*

Goldstein, H., & Scott, M. S. (2005). Shifting and sharing responsibility for public safety problems. *Problem-Oriented Policing Response Guide No. 3.* Retrieved

from http://www.popcenter.org/responses/responsibility/

Grabosky, P. N. (1995). Regulation by reward: On the use of incentives as regulatory instruments. *Law and Policy, 17*, 257–282. doi: 10.1111/j.1467-9930.1995.tb00150.x

Green Bay Police Department. (1999). *Street sweeping Broadway style: Revitalising a business district from the inside out.* Retrieved from http://www.popcenter.org/library/awards/goldstein/1999/99-22(W).pdf

Green, L. (1995). Policing places with drug problems: The multi-agency response team approach. In J. Eck and D. Weisburd (Eds.), *Crime Prevention Studies Vol. 4. Crime and place* (pp. 199–216). Monsey, NY: Criminal Justice Press.

Green, L. (1996). *Policing places with drug problems.* Thousand Oaks, CA: Sage.

Green-Mazerolle, L., & Roehl, J. (Eds.). (1998). *Crime Prevention Studies Vol. 9. Civil remedies and crime prevention.* Monsey, NJ: Criminal Justice Press.

Guerette, R. T., & Bowers, K. J. (2009). Assessing the extent of crime displacement and diffusion of benefits: a review of situational crime prevention evaluations. *Criminology, 47*, 1331–1368. doi: 10.1111/j.1745-9125.2009.00177.x

Hammerstrøm, K., Wade, A., & Jørgensen, A-M. K. (2010). *Searching for studies: A guide to information retrieval for Campbell systematic reviews.* doi: 10.4073/csrs.2010.1

Higgins, D. F., & Coldren, J. R. (2000). *Evaluating gang and drug house abatement in Chicago.* Chicago: Illinois Criminal Justice Authority.

Independent Police Commission. (2013). *Policing for a Better Britain* (Report). Retrieved from http://independentpolicecommission.org.uk/

Jennings, W. G., & Reingle, J. M. (2012). On the number and shape of developmental/life-course violence, aggression, and delinquency trajectories: A state-of-the-art review. *Journal of Criminal Justice, 40*, 472–489. doi: 10.1016/j.jcrimjus.2012.07.001

Jones, T., & Newburn, T. (Eds.). (2006). *Plural policing: A comparative perspective.* New York: Routledge.

Kelling, G., & Moore, M. (1988). From political reform to community: The evolving strategy of police. In J. R. Greene & S. Mastrofski, (Eds.), *Community policing: Rhetoric or reality?* (pp. 3 – 25). New York: Praeger.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Lipsky, M. (2010). *Street level bureaucracy: Dilemmas of the individual in public services* (2nd Ed.). New York: Russell Sage Foundation.

Littell, J. H., Corcoran, J., & Pillai, V. K. (2008). *Systematic reviews and meta-analysis*. New York: Oxford University Press.

Loader, I. (2000). Plural policing and democratic governance. *Social and Legal Studies, 9*, 323–345. doi: 10.1177/096466390000900301

Lum, C., Koper, C. S., & Telep, C. W. (2011). The Evidence-Based Policing Matrix. *Journal of Experimental Criminology, 7*, 3–26. doi: 10.1007/s11292-010-9108-2

Mason, M., & Bucke, T. (2002). Evaluating actions against local drug markets: A 'systematic' review of research. *Police Journal, 75*, 15–30.

Mazeika, D., Bartholomew, B., Distler, M., Thomas, K., Greenman, S., & Pratt, S. (2010). Trends in police research: A cross-sectional analysis of the 2000–2007 literature. *Police Practice and Research: An International Journal, 11*, 520–547. doi: 10.1080/15614263.2010.497390

Mazerolle, L. (2014). The power of policing partnerships: Sustaining the gains. *Journal of Experimental Criminology*. Advance online publication. doi: 10.1007/s11292-014-9202-y

Mazerolle, L., Kadleck, C., & Roehl, J. (1998). Controlling drug and disorder problems: The role of place managers. *Criminology, 36*, 371–404. doi: 10.1111/j.1745-9125.1998.tb01252.x

Mazerolle, L. G., Price, J. F., & Roehl, J. (2000). Civil remedies and drug control. *Evaluation Review, 24*, 212–241. doi: 10.1177/0193841X0002400203.

Mazerolle, L., & Ransley, J. (2005). *Third party policing*. Cambridge, UK: Cambridge University Press.

Mazerolle, L., & Ransley, J. (2006). The case for third party policing. In D. Weisburd & A. A. Braga (Eds.), *Police innovation: Contrasting perspectives* (pp. 191–206). Cambridge, UK: Cambridge University Press.

Meares, T. (2006). Third party policing: A critic. In D. Weisburd, & A. D. (Eds.), *Police innovation: Contrasting perspectives* (pp. 207–221). Cambridge: Cambridge University Press.

Meyer, S., & Mazerolle, L. (2014). Police-led partnership responses to high risk youths and their families: Challenges associated with forming successful and sustainable partnerships. *Policing and Society: An International Journal of Research and Policy, 24*, 242–260. doi: 10.1080/10439463.2013.784295

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behaviour: A

developmental taxonomy. *Psychological Review, 100*, 674–701. doi: 0.1037/0033-295X.100.4.674

Moffitt, T. E. (2003). Life-course persistent and adolescence-limited antisocial behaviour: A research review and a research agenda. In B. Lahey, T. E. Moffitt, & A. Caspi (Eds.), *The causes of conduct disorder and serious juvenile delinquency* (pp. 49–75). New York: Guilford.

Moher, D., Liberati, A., Altman, D. G., & The PRISMA Group. (2009). Preferred items for systematic reviews and meta-analysis: The Prisma Statement. *PLoS Med, 6*(7), e1000097. doi:10.1371/journal.pmed1000097

Neville, R., & Higginson, A. (2014). Optimal Search of Indexing Databases (OSID) (Computer software). Brisbane, Australia: University of Queensland.

Piquero, A. R. (2008). Taking stock of developmental trajectories of criminal activity over the life course. In A. M. Liberman (Ed.), *The long view of crime: A synthesis of longitudinal research* (pp. 23–78). New York: Springer.

Piquero, A. R., Jennings, W. G., & Barnes, J. C. (2012). Violence in criminal careers: A review of the literature from a developmental perspective. *Aggression and Violent Behaviour, 17*, 171–179. doi: 10.1016/j.avb.2012.02.008

Ransley, J. (2014). Legal frameworks for Third-Party Policing. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopaedia of Criminology and Criminal Justice* (pp. 2906–2914). New York: Springer.

Ransley, J., & Mazerolle, L. (2009). Policing in an era of uncertainty. *Police Practice and Research: An International Journal, 10*, 365–381.doi: 10.1080/15614260802586335

Ransley, J., Mazerolle, L., Manning, M., McGuffog, I., Drew, J. M., & Webster, J. (2011). *Reducing the methamphetamine problem in Australia: Evaluating innovative partnerships between police, pharmacies and other third parties* (Monograph Series No. 39). Canberra: National Drug Law Enforcement Research Fund. Retrieved from http://www.ndlerf.gov.au/publications/monographs/monograph-39

Raven, B. H. (2008). The bases of power and the power/interaction model of interpersonal influence. *Analysis of Social Issues and Public Policy, 8*, 1–22. doi: 10.1111/j.1530-2415.2008.00159.x

Rosenbaum, D. P. (2002). Evaluating multi-agency anti-crime partnerships: Theory, design and measurement issues. In N. Tilley (Ed.) *Crime Prevention Studies Vol. 14. Evaluation for crime prevention* (pp. 171–225). Monsey, NY: Criminal Justice Press.

Rosenbaum, D. P., & Schuck, A. M. (2012). Comprehensive community partnerships for

preventing crime. *Oxford Handbooks Online.* doi: 10.1093/oxfordhb/9780195398823.013.0012

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005*). Publication bias in meta-analysis: Prevention, assessment, and adjustments.* Chichester, UK: John Wiley & Sons.

Roussos, S. T., & Fawcett, S. B. (2000). A review of collaborative partnerships as a strategy for improving community health. *Annual Review of Public Health, 21*, 369–402. doi: 10.1146/annurev.publhealth.21.1.369

Scott, M. S. (2013). Policing for prevention: Shifting and sharing the responsibility to address public safety problems. In N. Tilley (Ed.), *A handbook for crime prevention: Theory, policy and practice (2nd ed., pp. 385–409).* London: Willan Publishing.

Shearing, C. (2007). Policing our future. In A. Henry & D. J. Smith (Eds.), *Transformations in of policing* (pp. 249–272). Aldershot: UK. Ashgate Publishing.

Sherman, L. W. (1992). Attacking crime: Police and crime control. In M. Tonry & N. Morris (Eds.), *Crime and justice: A review of research Vol. 15* (pp. 159–230). doi: 10.1086/449195.

Sherman, L. W. (2002). Fair and effective policing. In J. Q. Wilson & J. Petersilia (Eds.), *Crime: Public policies for crime control* (pp. 383–412). Oakland, CA: ICS Press.

Sherman, L. W. (2013). The rise of evidence-based policing: Targeting, testing, and tracking. *Crime and Justice, 42*, 377–451. doi: 10.1086/670819

Sherman, L. W., Farrington, D. P., Welsh, B. C., & MacKenzie, D. L. (Eds.) (2006). *Evidence-based crime prevention* (Rev. ed.). New York: Routledge.

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology, 27*, 27–56. doi: 10.1111/j.1745-9125.1989.tb00862.x

Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J. E., Reuter, P., & Bushway, S. D. (1997). *Preventing crime: What works, what doesn't, what's promising.* Washington, DC: U.S. Department of Justice, National Institute of Justice.

Skogan, W., & Frydl, K. (Eds.). (2004). *Fairness and effectiveness in policing: The evidence.* Washington, DC: National Academies Press.

Spelman, W., & Eck, J. E. (1987). *Problem-Oriented Policing.* Washington, DC: US Department of Justice.

Telep, C. W. (2009). Citation analysis of randomised experiments in criminology and criminal justice: A research note. *Journal of Experimental Criminology, 5, 441–463.*

doi: 10.1007/s11292-009-9082-8

Telep, C. W., Varriale, J. A., Gibbs, J. C., Chongmin, N., & Bartholomew, B. (2008). Trends in police research: A cross-sectional analysis of the 2005 literature. *Police Practice and Research: An International Journal, 9*, 445–469. doi: 10.1080/15614260802551610

Telep, C., & Weisburd, D. (2012). What is known about the effectiveness of police practices in reducing crime and disorder? *Police Quarterly, 15*, 331–358. doi: 10.1177/1098611112447611

Thornberry, T., Huizinga, D., & Loeber, R. (1995). The prevention of serious delinquency and violence: Implications from the program of research on the causes and correlates of delinquency. In J. B. Howell, B., Krisberg, & D. J. Hawkins (Eds.), *Source book on serious, chronic, and violent offenders* (pp. 213–237). Thousand Oaks: CA: Sage.

Townsley, M. (2008). Visualising space time patterns in crime: The hotspot plot. *Crime Patterns and Analysis, 1*(1), 61–74.

Tummers, L. (2011). Explaining the willingness of public professionals to implement new policies: A policy alienation framework. *Administrative Sciences, 77*, 555–581. doi: 10.1177/0020852311407364

Uittenbogaard, A., & Ceccato, V. (2012). Space-time clusters of crime in Stockholm, Sweden. *Review of European Studies, 4*, 148–156. doi: 10.5539/res.v4n5p148

van Koppen, P. J., & Jansen, R. W. J. (1999). The time to rob: Variations in time and number of commercial robberies. *Journal of Research in Crime and Delinquency, 36*, 7–29. doi: 10.1177/0022427899036001003

Varriale, J. A., Gibbs, J. C., Ahlin, E. M., Gugino, M. R., & Chongmin, N. (2007). Trends in police research: A cross-sectional analysis of the 2004 literature. *Police Practice and Research: An International Journal, 8*, 461–485. doi: 10.1080/15614260701764330

Weber, L. (2013). *Policing non-citizens.* Abingdon, UK: Routledge.

Weisburd D., Groff, E. R., & Yang S.-M. (2012). *The criminology of place: Street segments and our understanding of the crime problem.* New York: Oxford University Press.

Weisburd, D., & Braga, A. A. (Eds.). (2006). *Police innovation: Contrasting perspectives.* Cambridge: Cambridge University Press.

Weisburd, D., & Eck, J. (2004). What can police do to reduce crime, disorder, and fear? *Annals of the American Academy of Political and Social Science, 593*, 42–65. doi: 10.1177/0002716203262548

Weisburd, D., & Green, L. (1995). Policing drug hot spots: The Jersey City drug market

analysis experiment. *Justice Quarterly, 12*, 711–735. doi: 10.1080/07428829500096261

Weisburd, D., & Mazerolle, L. (2000). Crime and disorder in drug hot spots: Implications for theory and practice in policing. *Police Quarterly, 3*, 331–349. doi: 10.1177/1098611100003003006

Weisburd, D., Bushway, S., Lum, C., & Yang, S.-M. (2004). Trajectories of crime at places: A longitudinal study of street segments in the city of Seattle. *Criminology, 42*, 283–321. doi: 10.1111/j.1745-9125.2004.tb00521.x

Weisburd, D., Morris, N. A., & Groff, E. R. (2009). Hot spots of juvenile crime: A longitudinal study of arrest incidents at street segments in Seattle, Washington. *Journal of Quantitative Criminology, 25*, 443–467. doi: 10.1007/s10940-009-9075-9

Weisburd, D., Telep, C. W., Hinkle, J .C., & Eck, J. E. (2010). Is problem-oriented policing effective in reducing crime and disorder? *Criminology and Public Policy, 9*, 139–172. doi: 10.1111/j.1745-9133.2010.00617.x

Wolfgang, M. E. (1973). Crime in a birth cohort. *Proceedings of the American Philosophical Society, 117*, 404–411.

Zakocs, R. C., & Edwards, E. M. (2006). What explains community coalition effectiveness?: A review of the literature. *American journal of preventive medicine, 30*, 351-361. doi: 10.1016/j.amepre.2005.12.004

SOURCES OF SUPPORT

DECLARATIONS OF INTEREST

Professor Mazerolle is one of the founding TPP scholars, has evaluated a number of policing interventions with TPP components, and has published widely on TPP and related topics. Nevertheless, Professor Mazerolle is neither advocate nor critic of TPP; for example, chapter seven of her TPP book clearly articulates equity issues and potentially negative side effects of TPP. Professor Mazerolle is committed to generating a neutral and accurate review of the impact of TPP in order to make a meaningful contribution to crime control/prevention policy and practice, irrespective of whether the review's findings contradict her previous research endeavours.

REVIEW AUTHORS

Lead review author:

Name:	Lorraine Mazerolle
Title:	Professor
Affiliation:	The University of Queensland (UQ), School of Social Science
Address:	Room 440, Michie Building (9), UQ
City, State, Province or County:	St Lucia, Queensland
Postal Code:	4067
Country:	Australia

Phone:	+617 3346 7877
Email:	l.mazerolle@uq.edu.au

Co-author(s):

Name:	Angela Higginson
Title:	Dr
Affiliation:	Queensland University of Technology, School of Justice
Address:	Gardens Point Campus, 2 George Street
City, State, Province or County:	Brisbane, Queensland
Postal Code:	4000
Country:	Australia
Phone:	+61 7 3138 7130
Email:	angela.higginson@qut.edu.au

Name:	Elizabeth Eggins
Title:	Ms
Affiliation:	The University of Queensland (UQ), School of Social Science
Address:	Room 423, Michie Building (9), UQ
City, State, Province or County:	St Lucia, Queensland
Postal Code:	4067
Country:	Australia
Phone:	+617 3365 6307
Email:	e.eggins@uq.edu.au

ROLES AND RESPONSIBLIITIES

- Content: Mazerolle, Higginson, Eggins
- Systematic review methods: Higginson, Eggins
- Information Retrieval: Eggins (with additional research assistants)
- Statistical analysis: Higginson

PRELIMINARY TIMEFRAME

- Searches for published and unpublished studies — July 2014
- Staff training and piloting of eligibility and coding protocols — August 2014
- Relevance assessments for GPD — September 2014
- Relevance assessments for TPP review — November 2014
- Full-text coding and data extraction from eligible literature — December 2014
- Statistical analysis — January 2015
- Preparation of final report — February 2015

PLANS FOR UPDATING THE REVIEW

We plan to update the review every three years.

AUTHORS' RESPONSIBILITIES

By completing this form, you accept responsibility for preparing, maintaining and updating the review in accordance with Campbell Collaboration policy. The Campbell Collaboration will provide as much support as possible to assist with the preparation of the review.

A draft review must be submitted to the relevant Coordinating Group within two years of protocol publication. If drafts are not submitted before the agreed deadlines, or if we are unable to contact you for an extended period, the relevant Coordinating Group has the right to de-register the title or transfer the title to alternative authors. The Coordinating Group also has the right to de-register or transfer the title if it does not meet the standards of the Coordinating Group and/or the Campbell Collaboration.

You accept responsibility for maintaining the review in light of new evidence, comments and criticisms, and other developments, and updating the review at least once every five years, or, if requested, transferring responsibility for maintaining the review to others as agreed with the Coordinating Group.

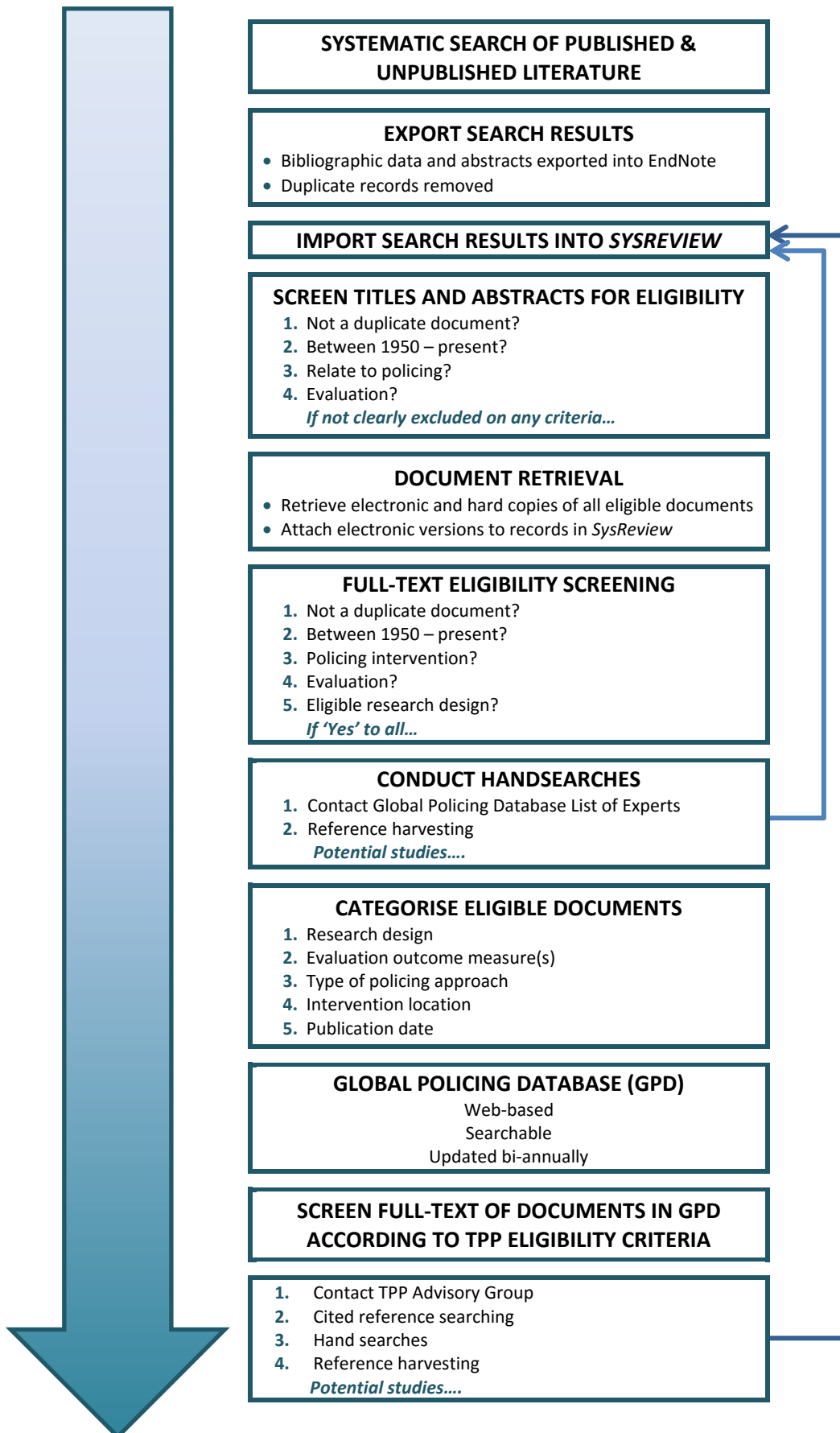## PUBLICATION IN THE CAMPBELL LIBRARY

The support of the Campbell Collaboration and the relevant Coordinating Group in preparing your review is conditional upon your agreement to publish the protocol, finished review and subsequent updates in the Campbell Library. Concurrent publication in other journals is encouraged. However, a Campbell systematic review should be published either before, or at the same time as, its publication in other journals. Authors should not publish Campbell reviews in journals before they are ready for publication in the Campbell Library. Authors should remember to include a statement mentioning the published Campbell review in any non-Campbell publications of the review.

I understand the commitment required to undertake a Campbell review, and agree to publish in the Campbell Library. Signed on behalf of the authors:

Form completed by: Angela Higginson

Date: 07 July 2014

APPENDIX A: GPD Flowchart

SYSTEMATIC SEARCH OF PUBLISHED & UNPUBLISHED LITERATURE

EXPORT SEARCH RESULTS
- Bibliographic data and abstracts exported into EndNote
- Duplicate records removed

IMPORT SEARCH RESULTS INTO *SYSREVIEW*

SCREEN TITLES AND ABSTRACTS FOR ELIGIBILITY
1. Not a duplicate document?
2. Between 1950 – present?
3. Relate to policing?
4. Evaluation?
   If not clearly excluded on any criteria…

DOCUMENT RETRIEVAL
- Retrieve electronic and hard copies of all eligible documents
- Attach electronic versions to records in *SysReview*

FULL-TEXT ELIGIBILITY SCREENING
1. Not a duplicate document?
2. Between 1950 – present?
3. Policing intervention?
4. Evaluation?
5. Eligible research design?
   If 'Yes' to all…

CONDUCT HANDSEARCHES
1. Contact Global Policing Database List of Experts
2. Reference harvesting
   Potential studies….

CATEGORISE ELIGIBLE DOCUMENTS
1. Research design
2. Evaluation outcome measure(s)
3. Type of policing approach
4. Intervention location
5. Publication date

GLOBAL POLICING DATABASE (GPD)
Web-based
Searchable
Updated bi-annually

SCREEN FULL-TEXT OF DOCUMENTS IN GPD ACCORDING TO TPP ELIGIBILITY CRITERIA

1. Contact TPP Advisory Group
2. Cited reference searching
3. Hand searches
4. Reference harvesting
   Potential studies….

APPENDIX B: GPD Search and Screening Methodology

The Global Policing Database (GPD) will be a web-based freely accessible and searchable database designed to capture all published and unpublished experimental and quasi-experimental evaluations of policing interventions published since 1950. To compile the GPD we will systematically search for, retrieve and screen published and unpublished literature that reports on impact evaluations of policing interventions from 1 January 1950. There will be no restrictions on the type of policing technique, type of outcome measure or language of the research and we plan to update the GPD biennially. Appendix A summarises the GPD methodology.

SEARCH STRATEGY FOR IDENTIFYING STUDIES

Search Locations

To reduce publication and discipline bias, our search strategy for the GPD will have an international scope and involve searching for published and unpublished literature across a number of disciplines (e.g., criminology, law, political science, public health, sociology, social science and social work). We will capture a comprehensive range of published (i.e., journal articles, book chapters) and unpublished literature (e.g., working papers, governmental reports, technical reports, conference proceedings, dissertations) by implementing a search strategy with four stages:

1. Searching bibliographic, grey literature, and dissertation databases
2. Searching relevant websites
3. Reference harvesting of eligible studies and previous reviews
4. Contacting policing experts and authors of eligible studies for feedback and input

Table 1 lists the search locations that will be used in each of these stages. We have identified a wide range of non-English search locations to ensure the GPD has an international scope, and we will translate the search terms when a search in English is not appropriate. We will consult with our Methods Advisory Group to identify additional foreign language search locations not already captured by our search strategy. We have invited a number of systematic search and information specialists to form a Methods Advisory Group. Invitees include: Professor Mark Lipsey (Peabody Research Institute, Vanderbilt University), Phyllis Shultze (Don M. Gottfredson Library of Criminal Justice, Trial Search Coordinator for Campbell Collaboration Crime and Justice Group), Jon Eyers (information specialist for the Campbell Collaboration International Development Group), and members of the World Criminal Justice Library Network (http://andromeda.rutgers.edu/~wcjlen/WCJ/index.htm).

Table 1. GPD Search Locations

INDEXED & ACADEMIC DATABASES	

ProQuest	Criminal Justice
	Dissertation and Theses Database (Social Sciences)
	Political Science
	Periodical Archive Online
	Research Library (Social Science)
	Social Science Journals
	Sociology
	CSA Illumina
	    Applied Social Sciences Index and Abstracts (ASSIA)
	    International Bibliography of the Social Sciences
	    Public Affairs Information Service (PAIS International)
	    Social Services Abstracts
	    Sociological Abstracts
	    Worldwide Political Sciences Abstracts

EBSCO	Academic Search Premier Criminal Justice Abstracts EconLit MEDLINE with Full-Text Social Sciences Full-Text
OVID	International Political Science Abstracts (IPSA) PsycARTICLES PsycEXTRA PsycInfo Social Work Abstracts

Web of Knowledge	Current Contents – Social and Behavioural Sciences Edition
	Web of Science
	    Book Citation Index (Social Sciences and Humanities)
	    Conference Proceedings Citation Index (Social Sciences and Humanities)
	    Humanities
	    Social Science Citation Index

Informit	AGIS Plus Text Australian Criminology Database Australian Federal Police Database Australian Public Affairs Information Service Full-Text DRUG Health & Society Database Humanities and Social Sciences Collection (Law, Social Sciences subsets)
Gale-Cengage	Expanded Academic ASAP
Standalone Databases	African Journals Online Alcohol and Alcohol Problems Science Database American Bibliography of Slavic & Eastern European Studies (ABSEES) C2-SPECTR Cairn (French) California Commission on Peace Officer Standards & Training (POST) Library Cambridge Journals Online Central and Eastern European Online Library (CEEOL)

	China Academic Journals (incl. China Doctoral Dissertations) Chinese Social Sciences Citation Index Clase (Spanish and Portuguese) Cochrane Library CrimDoc (Canada) CrimeSolutions.gov Directory of Open Access Journals Drug Policy Alliance – Lindesmith Library DrugScope E-Korean Studies Database E-Library.ru (includes Russian Science Citation Index) Evidence-Based Policing Matrix FBI – The Vault FORENSICnetBASE HeinOnline Index Islamicus Indian Citation Index (Social Science and Humanities subset) International Initiative for Impact Evaluation Database (3ie) Japanese Periodical Index JSTOR LawENFORCEMENTnetBASE LILACS (Spanish and Portuguese) MultiData Online (Index to Arabic Periodicals) National Criminal Justice Reference Service OCLC FirstSearch (WorldCat) Oxford Scholarship Online Periodica (Spanish and Portuguese) Persee (French) RCT Documentation Centre and Library RefDoc (French) Russian Academy of Sciences Bibliographies SafetyLit Database Sage Journals Online and Archive (Sage Premier) SAGE Knowledge SciELO (Spanish and Portuguese) ScienceDirect SCOPUS SpringerLink Taylor & Francis Online Universal Database of Russian Social Sciences and Humanities Publications Wiley Online Library YU-DSpace Repository (Arabic)
GREY LITERATURE **SOURCES & WEBSITES**	American Institutes for Research Australian Institute of Criminology Brå Brottsforebyggande radet (Swedish National Council for Crime Prevention) Bureau of Police Research and Development (India) Bibliography of Nordic Criminology Canadian Evaluation Society Canadian Police Research Centre Canadian Policy Research Networks Center for Evidence-Based Crime Policy Center for Problem-Oriented Policing Centre for Crime Prevention in Lithuania (CCPL) Centre for Criminology (China)

Centre for Excellence in Policing and Security (CEPS, Australia)
Centre of Criminology (South Africa)
College of Policing (including POLKA, UK)
Crime and Justice Research Centre (New Zealand)
CrimPrev.dk (Danish)
Crime Research Centre (Western Australia)
Current Social Science Research Reports (CSSRR)
Economic and Social Research Council (EBSRC)
European Crime Prevention Network
European Institute for Crime Prevention and Control
European Police College (CEPOL)
Evidence for Policy and Practice Information and Coordinating Centre (EPPI-Centre)
Gray Literature Database (Don M. Gottfredson Library of Criminal Justice)
GreySource
Homeland Security Digital Library (HSDL)
Home Office (UK)
Institute for Security Studies (South Africa)
Institute for Criminal Policy Research (UK)
Institute of Criminology and Social Prevention (Czech Republic)
Jill Dando Institute of Crime Science (JDI)
Key Centre for Ethics, Law, Justice and Governance (Griffith University, Australia)
Kriminologisches Forschunginstitut Niedersachsen (Criminological Research Institute, Germany)
Kriminologische Zentralstelle (German Centre for Criminology)
Kriminologiska Institutionen DiVA (Stockholm University Department of Criminology Digital Scientific Archive)
National Clearinghouse for Science, Technology, and the Law
National Crime Prevention Council of Singapore
National Institute for Research Advancement Policy Research Watch database
National Institute of Criminology (Hungary)
National Institute of Justice (NIJ, US)
National Registry of Evidence-Based Program and Practice (NREPP)
National Research Institute of Police Science (Japanese)
National Technical Information Service (NTIS)
NSW Bureau of Crime Statistics and Research (BOSCAR)
OAIster
Office of Community Oriented Policing Services (COPS)
Office of Juvenile Justice and Delinquency Prevention (US)
Open System for Information on Grey Literature In Europe (OpenGrey)
Organisation for Economic Cooperation and Development Library (OECD)
Police Executive Research Forum (US)
Police Foundation (US)
Policing Online Information System (POLIS, Europe)
ProjectCork.org
RAND Corporation Research Services
Russian Eurasian Security Network (RES)
Scandinavian Research Council for Criminology
Scottish Centre for Criminology
Scottish Institute for Policing Research
Social Science Research Network
South Australian Office of Crime Statistics and Research (OSCAR)
Tasmania Institute of Law Enforcement Studies (TILES)
United Nations Educational, Scientific and Cultural Organisation (UNESDOC)
United Nations Interregional Crime and Justice Research Institute (Documentation

	and Information Centre, UNICRI) United Nations Office on Drugs and Crime (UNODC) WODC Internet Sources Guide WorldBank World Criminal Justice Library Network 'Criminal Justice Links Annotated' (all categories of sources will be searched across all the countries listed on the website; n = 102 countries and 'International')[5].
HANDSEARCHES OF PREVIOUS REVIEWS & BIBLIOGRAPHIES	Bartholomew et al. (2009) Beckman et al. (2003, 2005) Braga et al. (2014) Braga & Weisburd (2007) Farrington (1983) Farrington & Welsh (2005, 2006) Gibbs et al. (2006) Mason & Bucke (2002) Mazeika et al. (2010) Mazerolle & Ransley (2005) Michigan State University Sherman (1992, 2002, 2013) Sherman et al. (1997, 2006) Skogan & Frydl (2004) Telep (2009) Telep et al. (2008) Telep & Weisburd (2012) Varriale et al. (2007) Weisburd & Eck (2004) World Criminal Justice Library Network bibliographies in the 'Police and Law Enforcement' category (http://andromeda.rutgers.edu/~wcjlen/WCJ/mainpages/bibliogs.htm)

Our search strategy includes the following languages:

- **African**
- **Arabic**
- **Chinese**
- **English**
- **French**
- **German**
- **Hindi**
- **Japanese**
- **Korean**
- **Portuguese**
- **Russian**
- **Spanish**

The search locations will be as exhaustive as possible; however, we note that there is substantial overlap of content coverage between many of the databases. Therefore, we will use the *Optimal Searching of Indexing Databases* (OSID) computer program (Neville & Higginson, 2014) to analyse the content crossover for all databases that have accessible content coverage lists. OSID will analyse and create a search location solution that provides the most comprehensive coverage via the least number of databases. For example, if the content for the set of databases seen in Figure 1 were imported, OSID would provide a solution that entails searching only databases 3 and 4 because the content covered by

---

[5] We will exclude the following categories from our search: Corrections, Human Rights, Law and the Courts.
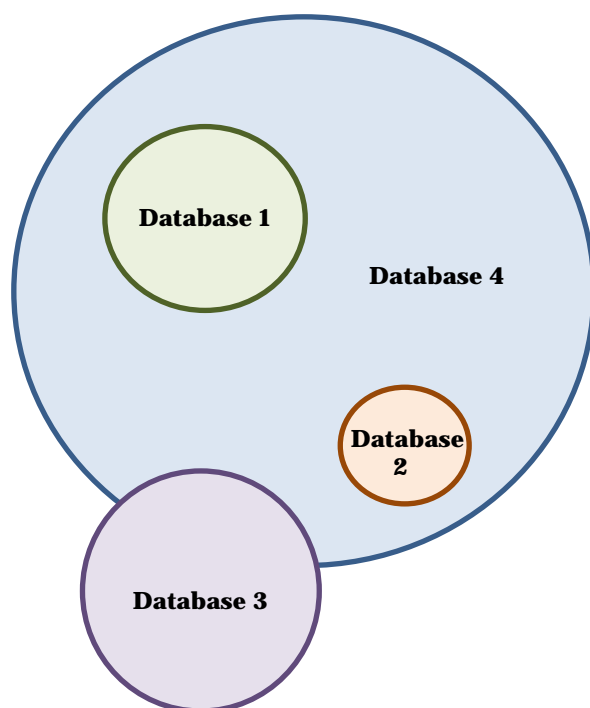
databases 1 and 2 is covered by database 4.

Figure 1. Example database content cross-over analysed by OSID

We will contact policing experts and authors of eligible studies after all documents have been screened for eligibility, to identify any eligible studies not captured in our search. Our group of policing experts is largely drawn from the new American Society of Criminology Division of Policing. All manually added studies will undergo the same screening and coding process as those retrieved from the systematic search (see Appendix A).

Search Terms

To ensure optimum sensitivity and specificity, our search strategy will utilise a combination of free-text and controlled vocabulary search terms. Because controlled vocabularies and search capabilities vary across databases, the exact combination of search terms and field codes will need to be adapted to each unique database. We will consult closely with the our Methods Advisory Group when devising the search strategies for each location. Search strategies for each search location will be reported in the final TPP review as per the guidelines provided in Campbell Collaboration information retrieval guide (Hammerstrøm, Wade, & Jørgensen, 2010).

The free-text search terms we will use for the GPD are provided in Table 2 and are grouped by intervention (i.e., some form of policing) and evaluation terminology. Although the search strategy across search locations will be unique, we will follow a number of general rules:

- Search terms will be combined into search strings using Boolean operators "AND" and "OR". Specifically, terms within each category will be combined with "OR" and

categories will be combined with "AND". For example: (police OR policing OR "law#enforcement") AND (analy* OR ANCOVA OR ANOVA* OR …).

- Compound terms (e.g., law enforcement) will be considered single terms in search strings by using quotation marks (i.e., "law*enforcement") to ensure that the database searches for the entire term rather than separate words.
- Wild cards and truncation codes will be used for search terms with multiple iterations from a stem word (e.g., evaluation, evaluate) or spelling variations (e.g., evaluat* or randomi#e).
- If a database has a controlled vocabulary term that is equivalent to "POLICE", we will combine the term in a search string that includes both the policing and evaluation free-text search terms. This approach will ensure that we retrieve documents that do not use policing terms in the title/abstract but have been indexed as being related to policing in the database. An example of this approach is the following search string: (((SU: "POLICE") OR (TI,AB,KW: police OR policing OR "law*enforcement")) AND (TI,AB,KW: intervention* OR evaluat* OR compar* OR …)).
- For search locations with limited search functionality, we will implement a broad search that uses only the policing free-text terms.
- Multidisciplinary database searches will be limited to relevant disciplines (e.g., include social sciences but exclude physical sciences).
- Search results will be refined to exclude specific types of documents that are not suitable for systematic reviews (e.g., newspapers, front/back matter, book reviews).

Table 2. Free-text search terms for the GPD systematic search

Policing Search Terms		Evaluation Search Terms	
police	analy*	"explanatory variable*"	pretest
policing	ANCOVA	hypothes*	"propensity score*"
"law*enforcement"	ANOVA	impact*	quantitative
constab*	"ABAB design"	"independent variable*"	"quasi#experiment*"
detective *	"AB design"	intervent*	questionnaire
sheriff*	baseline	interview*	random*
	causa*	longitudinal	RCT
	"chi#square*"	MANCOVA	regress*
	"comparison condition*"	MANOVA	result*
	"comparison group*"	"matched group"	"risk#ratio*"
	"control condition*"	measure*	sampl*
	"control group*"	"meta-analysis"	"standard deviation*"
	correlat*	"meta analysis"	statistic*
	covariat*	"odds#ratio*"	studies
	"cross#section*"	"outcome variable*"	study
	data	outcome*	survey*
	"dependent variable*"	paramet*	"systematic review*"
	effect*	"post-test"	"t#test*"
	efficacy	posttest	"time#series"
	eval*	predict*	treatment*
	experiment*	"pre-test"	variance

CRITERIA FOR INCLUDING STUDIES IN THE DATABASE

Each document must satisfy all inclusion criteria to be included in the GPD: timeframe, intervention and research design. There will be no restrictions applied to the types of outcomes, participants, settings or languages considered eligible for inclusion in the GPD.

Research timeframe

Because the 'reform' era of policing began in the 1960s (see Kelling & Moore, 1988; Ransley & Mazerolle, 2009), we anticipate that policing research will begin around this time period and increase over time to the present day. We have erred on the side of caution and will include research conducted after 1 January 1950.

Types of interventions

Each document must contain an impact evaluation of a policing intervention. We define a policing intervention is some kind of a strategy, technique, approach, activity, campaign, training, directive, or funding / organisational change that involves police in some way (other agencies or organisations can be involved). Police involvement is broadly defined as:

- Police initiation, development or leadership
- Police are recipients of the intervention or the intervention is related, focused or targeted to police practices
- Delivery or implementation of the intervention by police

Possible examples include: hot spots policing, third party policing, problem-oriented policing, legitimacy policing interventions, police investigative techniques, training programs for police recruits, interventions to reform policing organisations, interventions for managing human resources in policing.

Types of study design

The GPD will include quantitative research that uses randomised experimental (e.g., RCTs) and quasi-experimental evaluation designs with a valid comparison group that does not receive the intervention. We will accept designs where the comparison group receives 'business-as-usual' policing, no intervention or an alternative intervention (treatment-treatment designs).

Although not as robust as randomised experimental designs, 'strong' quasi-experiments can be used to provide causal inference when the nature of the design attempts to minimise threats to internal validity (see Farrington, 2003; Shadish, Cook, & Campbell, 2002). This can be achieved in a number of ways, such as: controlling the assignment of cases to treatment and control groups (regression discontinuity), matching the characteristics of the treatment and control groups (matched control), statistically accounting for differences between the treatment and control groups (designs using multiple regression analysis), or providing a difference-in-difference analysis (parallel cohorts with pre-test and post-test

measures). Therefore, we will include the following 'strong' quasi-experimental designs in the GPD:

- Regression discontinuity designs
- Designs using multivariate controls (e.g., multiple regression)
- Matched control group designs with or without pre-intervention baseline measures (propensity or statistically matched)
- Unmatched control group designs with pre-intervention measures (difference-in-difference analysis)
- Short interrupted time-series designs with control group (less than 25 pre- and 25 post-intervention observations (Glass, 1997))
- Long interrupted time-series designs with or without a control group (≥25 pre- and post-intervention observations (Glass, 1997))

A third group of research designs, 'weak quasi-experiments', will be included in the GPD. Although not as reliable as experiments or strong quasi-experiments for demonstrating causality, 'weak' quasi-experiments can be used to demonstrate the magnitude of the relationship between an intervention and an outcome. Therefore, we will include the following 'weak' quasi-experimental designs in the GPD:

- Unmatched control group designs with pre-post intervention measures which allow for difference-in-difference analysis
- Unmatched control group designs without pre-intervention measures where the control group has face validity
- Raw unadjusted correlational designs where the variation in the level of the intervention is compared to the variation in the level of the outcome
- Treatment-treatment designs[6]

We will exclude single group designs with pre- and post-intervention measures as these designs are highly subject to bias and threats to internal validity.

SCREENING AND CODING STAGES

Title and abstract screening

We will export the full search results from EndNote (duplicates removed) into *SysReview*, a Microsoft Access database for screening and coding research that is customisable to individual review requirements (see Appendix C for screen shots). The title and abstract of each document will be screened by trained research staff, using the screening companion in

---

[6] Whilst we acknowledge this design can be methodologically robust (e.g., units of analysis are randomly assigned to treatments), this type of design generally provides indications of the comparative effectiveness of different interventions rather than providing indications of causality.

Appendix D, to identify potentially eligible research that satisfy the following criteria:

- Document is dated between 1950 – present
- Document is unique
- Document is related to policing
- Document is an impact evaluation

Documents will be excluded if the answer to any one of the criteria is unambiguously 'No', and will be classified as potentially eligible otherwise. We will err on the side of inclusivity and only exclude studies where it is clear that these criteria are not met. Documents classified as potentially eligible will progress to the full-text eligibility screening stage.

Full-text eligibility screening

Wherever possible, a full-text electronic version of eligible records will be imported into SysReview. For records without an electronic version, a hardcopy of the record will be located to enable full-text eligibility screening. The full-text of each document will be screened in two stages, using the screening companion in Appendix E, to identify studies that satisfy the following criteria:

- Document is dated between 1950 – present
- Document is unique
- Document is reports on a policing evaluation
- Document reports on a quantitative impact evaluation
- Evaluation uses an eligible research design

Documents that are not excluded during either of the screening stages will progress to the in-depth coding phase. To ensure consistency in screening decisions, each document coder will screen 30 documents for eligibility and inter-coder agreement will be calculated (percentage agreement between coders that document is eligible). We will accept an inter-coder agreement of 95 percent or better. If there is less than 95 percent agreement, we will implement further training and rescreen the group of documents where agreement fell below the 95 percent threshold. Disagreements regarding the eligibility of training and non-training documents will be resolved by a discussion between the coders and the review manager.

After the eligibility screening phase has been completed, a list of eligible documents and the inclusion criteria will be distributed to the policing experts for perusal to ensure that eligible studies have not been omitted from the review. Any additional studies will be assessed for eligibility in the same manner as studies retrieved from the systematic search.

Full-text coding

A team of trained research assistants will code the documents using the coding companion in Appendix F. Documents will be read in detail and coded according to:

- Publication date of the document
- Language of the document
- Location of the intervention
- Type of problem targeted by the intervention
- Type of outcome measure(s) used to evaluate the intervention
- Type of participants used to evaluate the intervention
- Type of policing intervention evaluated

Each document may (a) report multiple outcomes for the one intervention or (b) contain multiple studies with multiple outcomes. SysReview allows for this nested data situation by enabling coders to add multiple outcomes for each unique study, and to add multiple studies within the one document record. The results of the eligibility screening and coding phases will be presented in the final review in the form of a PRISMA flowchart (Moher et al., 2009).

We will assess coders' understanding of the coding structure and consistency of coding decisions by implementing the same quality control process used for the eligibility screening phase. If there is missing data for key coding fields in the original document, we will attempt to correspond with the document's author(s) to obtain the required information.

APPENDIX C: Full-Text Eligibility Screening Companion for TPP Review

Use this document together with the review protocol to help with completing eligibility screening.

Before Screening

1. Open the 'TPP Review' Access database.

2. Click on '*Eligibility screening*' in the list on the left hand side of the screen.

3. Select your name from the '*User*' drop down menu and then click on the '*Go to First Unscreened Title*' button.

4. The form that will appear is divided into two parts.

    a. The top section contains the following document information:

        • **TitleID.** This the unique identification number for this document.

        • **Full citation.** A full reference in APA format (6th Edition) should be present.

        • **Document attachment.** If the document has been electronically attached, there will be a PDF or Microsoft Word document icon. Double-click on the icon and then double-click on the attachment in the dialogue box to open the document.

    b. The bottom part of the form contains inclusion/exclusion criteria that are used to determine the document's eligibility for the review. Eligible documents will proceed to the more in-depth coding stage. If a document is eligible, the text above the '*Complete Screening*' button at the bottom of the form will read '*Title is eligible*'. If a document is not eligible for inclusion in the review, it will read '*Title is NOT eligible*'.

5. When you have finished screening the document, click the '*Complete Screening*' button at the bottom of the form. Your name and today's date should appear beside '*Screened by*'. To move to the next document, click on the 'Go to First Unscreened Title' button at the top of the form.

Screening Criteria

Please read the document in enough detail to be able to address the inclusion and exclusion criteria **with certainty**. If you are unsure at any stage, please consult with the review manager to ensure you make the most accurate screening decision. If a document needs to be excluded, you will need to click on the relevant criteria. The criteria will then become highlighted and you will not need to screen the document any further.

Criteria	Information
Document is not dated between 1980-2013	If the document is dated before 1980 or contains research that was conducted prior to 1980, select this criteria and you will not need to screen the document any further.
The intervention does not involve one or more third parties	If the document does not refer to one or more third parties and their involvement in an intervention, select this criteria. A third party must be an entity, individual or network of collaborating

Criteria	Information
	agencies that are external to police. For example: property owner, business owner, regulatory authority (e.g., liquor licensing, local council), government department (e.g., education department).
The intervention does not involve one or more legal levers	If the document does not refer to one or more legal levers, select this criteria. Legal levers are the legal powers possessed by third parties. For more detailed information, please see your training materials, the protocol or Table 4.2 in Mazerolle and Ransley (2005).
Components do not combine to form a TPP intervention	Select this criteria if police, third parties and legal lever(s) do not combine to form a TPP intervention. In a TPP intervention, police work in partnership with one or more third parties in order to access and utilise the third parties' legal lever(s) to control or prevent a crime problem.
Document does not report on a crime or disorder outcome	If the document does not report at least one crime or disorder outcome, select this criteria. All types of measures of crime or disorder are eligible for this review. **Note:** fear of crime is not a measure of crime or disorder. Sources of data for crime and/or disorder outcome variables include: - Official measures of crime (e.g., crime rates, calls-for-service, court outcomes data) - Unofficial measures of crime (e.g., citizen, police or third party perceptions collected by interview or survey methods) - Diffusion of crime control benefits - Displacement of crime and/or disorder - Systematic observation of physical disorder (e.g., graffiti, vandalism, abandoned buildings, dilapidated buildings, rubbish etc.) - Systematic observation of social disorder (loitering, solicitation, drunk/disorderly conduct in public places, excess noise etc.)

APPENDIX D: SysReview Screen Shots

Figure D.1. Example Title and Abstracting Screening in SysReview
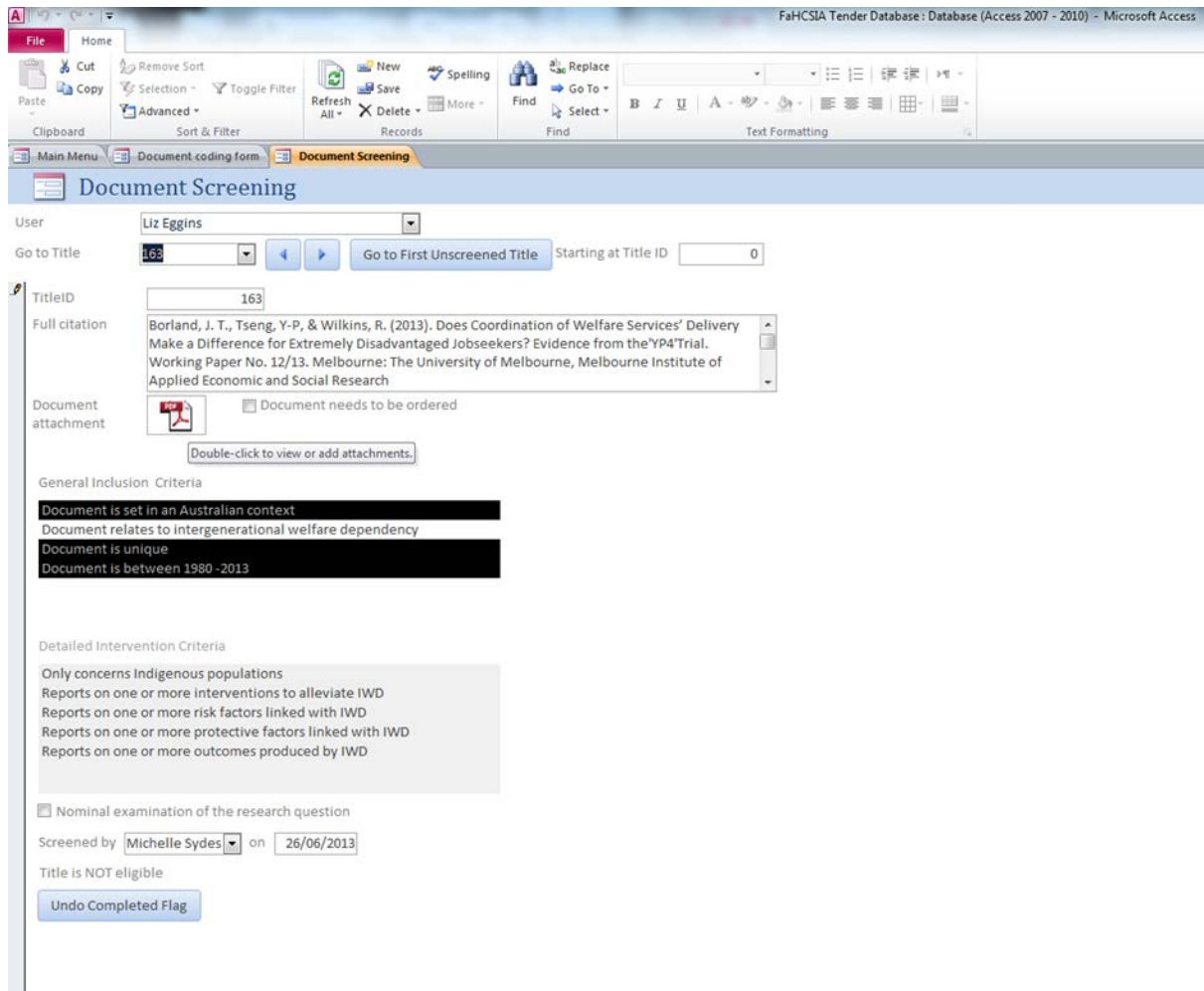
Figure D.2. Example Eligibility Screening in SysReview

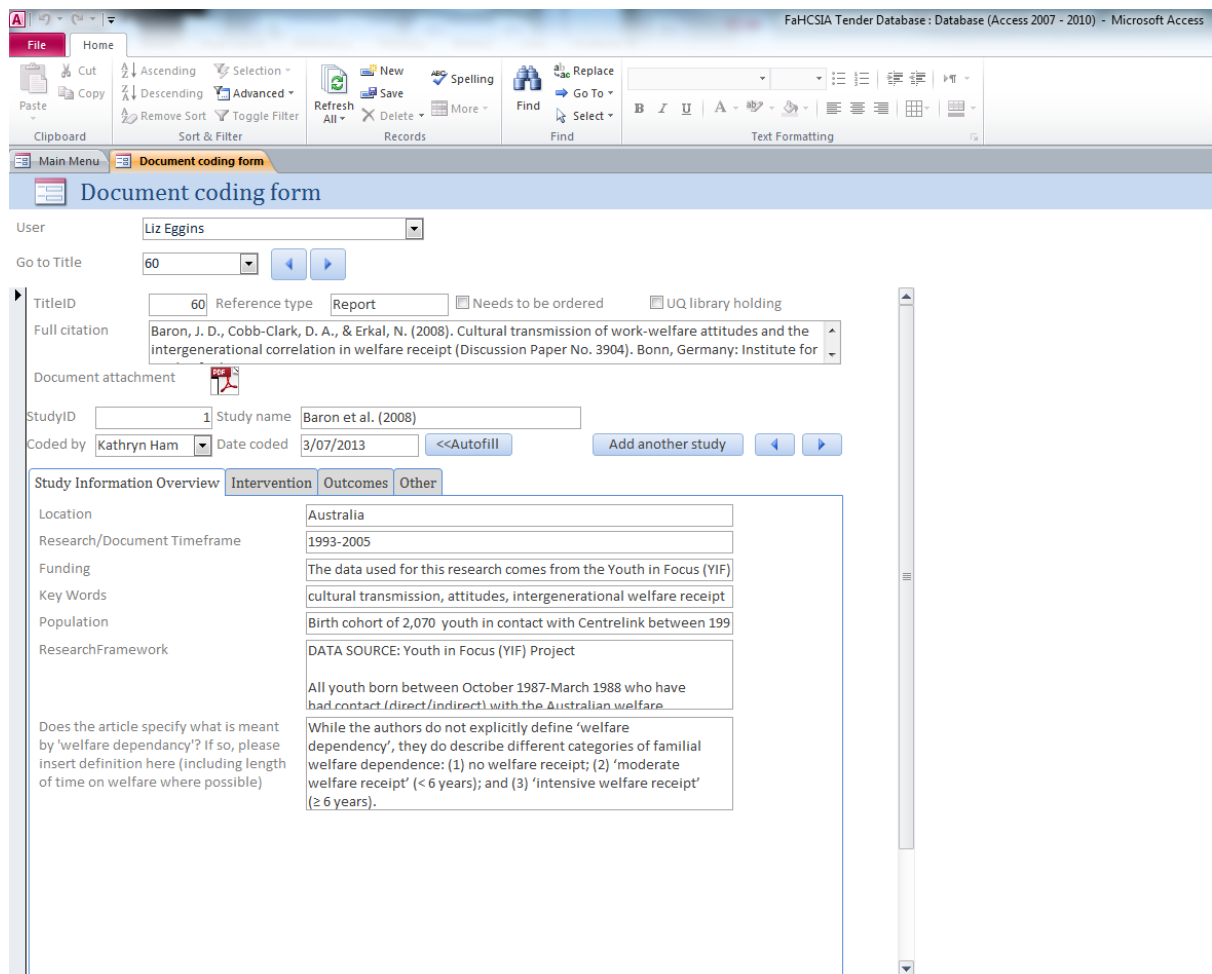**Figure D.3.** Example Full-Text Coding in SysReview

APPENDIX E: TPP Advisory Group Members

Member	Affiliation
Professor Anthony Braga	School of Criminal Justice, Rutgers University
Dr Russell Brewer	Centre of Excellence in Policing and Security, Griffith University
Professor Michael Buerger	Criminal Justice Program, Bowling Green State University
Dr Adrian Cherney	Humanities and Social Science (Criminology), University of Queensland
Assistant Professor Nicholas Corsaro	School of Criminal Justice, University of Cincinnati
Professor Benoit Dupont	School of Criminology, University of Montreal
Professor John Eck	Center for Criminal Justice Research, University of Cincinnati
Professor Ian Loader	All Souls College, Oxford University
Mr Peter Neyroud	Centre for Criminology, Cambridge University
Associate Professor Janet Ransley	School of Criminology and Criminal Justice, Griffith University
Dr Jan Roehl	Justice Research Center, California

APPENDIX F: Full-Text Coding Companion for TPP Review

Use this document together with the review protocol to help with completing title and abstract screening.

Before Coding

1. Open the 'TPP Review' SysReview database.

2. Select your name from the '*User*' drop down

3. Click on '*Code eligible documents'* in the list on the left hand side of the screen.

4. Select your user name from the top box and then click on the '*Go to First Uncoded Title*' button.

5. The form that will appear is divided into three parts.

    a. The top section contains the following document information:

       • **TitleID.** This the unique identification number for this document.

       • **Reference type.** This indicates if it is a book, journal paper etc.

       • **Needs to be ordered/UQ library holding checkboxes.** These checkboxes indicate whether the document was ordered in.

       • **Citation Fields.** Reference Type (e.g., Journal Article, Book Chapter etc), full citation in APA 6th Edition format.

       • **Document attachment.** A PDF version of the document should be accessible here. Double-click on the icon and then double-click on the attachment in the dialogue box to open the document.

    b. The second section contains information for each study in the document:

       • **Study ID.** Enter the first author (followed by et al. if >1 author), year of publication and name of the intervention (e.g., Brown et al. (2005)_SMART). If there is no name provided, enter the first author (followed by et al. if >1 author), year of publication and intervention location (e.g., Brown et al. (2005)_California).

       • **Study name.** Enter the name of the study if one is provided in the document. If there is no study name provided, enter the location where the study was conducted.

       • **Person coding and coding date.** Click on the '*<<Autofill*' button and your name and today's date should appear in the '*Coded by*' and '*Date coded*' boxes.

       • **Add another study.** Some documents may contain multiple studies. After you code the first study, click this button to code an additional study. You must code each study in full. You can scroll between the studies for an individual record by using the arrow buttons adjacent to the '*Add another study*' button. Separate studies are those that involve different interventions, or those where the same intervention is delivered at geographically or temporally distinct sites with their own control groups. Carefully consider the question of independence to determine whether it is a separate study, including issues of displacement and diffusion of benefit. Note: a document containing multiple studies is different from a document containing multiple outcomes. If you are unsure if a document contains multiple studies or multiple outcomes, please discuss the document with the review manager.

c. The bottom part of the form contains several tabs where you will need to record information about the intervention and the evaluation. This information is extracted from the attached document. If you are required to type information, either paraphrase from the document or place quotation marks around copy-and-pasted text so that we do not accidentally plagiarise when compiling the review report.

Coding Criteria

Please read the full-text of the document in enough detail to be able to complete all the forms **with certainty**. If you are uncertain about a coding decision, please discuss the issue with the review manager so that consistent and accurate coding decisions are made.

Do not leave any coding fields blank. If you cannot find the information in the document, do not leave the question unanswered. Either select 'Unsure' from the list of options or write 'Not specified' so that we know the information for the question is missing from the document rather than missed during coding.

Study Details Tab

Coding Field	Information
Intervention location	Enter as much detail as possible for the location of the TPP intervention. For example, country, state and/or city. Do not confuse this with the country where the document was published or the author's location.
Document language	Type in the language that the document was published in. If the document has been translated, type the language of the original document.
Research timeframe	Type in when the intervention started in the first box and when the intervention finished in the second box. The evaluation does not form part of the intervention timeframe.
Funding source	Using the boxes provided, specify who funded the (a) intervention, and (b) evaluation.
Intervention evaluator	Select who evaluated the intervention from the list provided: a. Academic scholars/institution b. Non-governmental organisation c. Government agency d. Police e. Other (specify)

Intervention Tab

This tab contains a number of coding fields relating to the characteristics of the TPP intervention (i.e. the treatment condition).

Focus of the Intervention

Coding Field	Information
Explicitly TPP	Did the study explicitly refer to the intervention as a TPP intervention? Select 'Yes' or 'No'. If 'No', use the box provided to enter what the author(s) called the intervention.
Exclusively TPP	Was the TPP intervention part of another policing intervention? Select 'Yes' or 'No'. If 'Yes', use the box provided to enter the type of policing intervention.
Control or prevention	What was the aim of the intervention? a. Select 'Control' if the intervention aimed to control an existing crime and/or disorder problem. b. Select 'Prevention' if the intervention aimed to prevent a future crime and/or disorder problem. c. Select 'Unsure' if the document does not provide enough information to answer this question.
Police-led	Select 'Yes' if police led the intervention. Generally, an intervention is police-led if the intervention description explicitly states or suggests that police coordinated the intervention (e.g., the police are primarily responsible for the intervention) Select 'No' if the intervention description identifies someone else as the study leader or coordinator. Select 'Unsure' if there is not enough information in the document to ascertain clear intervention leaders.

TPP Components: Third Parties, Legal Levers, Partnership and Author Conclusions

Coding Field	Information
Number of third parties [*dropdown menu*]	Use the dropdown menu to specify the number of third parties present in the intervention: a. Single third party b. Multiple third parties **Remember:** a third party MUST have a legal lever.
Type(s) of third party	Using the checkboxes provided to select the type(s) of third party(s)

[*Checkboxes*]	present in the intervention. Select all options that apply:
	
	   a. Individual third party
	   b. Organisational third party
	   c. Business third party
	   d. Regulatory authority third party
	   e. Government department third party
	
	Remember: a third party MUST have a legal lever.

Legal lever description [*Textbox*]	Use the textbox provided to describe the legal lever(s) used in the intervention. Provide as much detail as possible and note the page number of the document where the legal lever information is reported.

Type of engagement techniques
 	
 [*dropdown menu*]
	Specify the way that the police engaged with third parties during the intervention.
	
	1. Collaborative engagement techniques
	   a. Yes
	   b. No
	   c. Unsure
	
	2. Coercive engagement techniques
	   a. Yes
	   b. No
	   c. Unsure

Author conclusions	What did the author(s) conclude about the TPP intervention?
	
	   a. Intervention reduced crime or disorder
	   b. Intervention increased crime or disorder
	   c. Intervention had no effect on crime or disorder
	   d. Unclear / no conclusion stated by authors

Methodology Tab

Because all documents in the TPP database have already been categorised by research design, the first coding field, 'Research design' will be completed for you. The remaining fields ask for additional information about the research design used to evacuate the TPP intervention.

Coding Field	Information
Allocation / assignment process [*dropdown menu with textbox*]	How were units of analysis allocated or assigned to the experimental conditions / groups? a. Simple random assignment b. Random allocation in pairs, blocks or other randomised technique c. Haphazard assignment d. Adjacent geographic area (e.g., natural experiment) e. No control group (e.g., interrupted time-series or cohort designs) f. Other (please specify in adjacent text box)
Matching process [*dropdown menu with textbox*]	How were the experimental and comparison groups matched? a. Statistically matched b. Matched on administrative data (e.g., number of calls-for-service) c. Propensity score matching d. No matching e. Other (please specify in adjacent text box)
Matching variables [*textbox*]	Using the text box provided, enter the variables used for matching and the page number where the matching variables are listed in the document. Enter 'No matching' in the text box if the treatment and control conditions were not matched.
Unit of analysis [*dropdown menu with textbox*]	Specify the unit of analysis for the experimental <u>and</u> comparison conditions (where there is one) by selecting one of the following options: a. Individuals b. Groups of individuals (e.g., gangs) c. Geographic areas d. Multiple units of analysis e. Other (specify in adjacent text box)
Type of comparison condition [*dropdown menu*]	Specify the type of comparison condition used in the evaluation: a. No treatment b. "Treatment-as-usual" / "Business-as-usual" c. An alternative treatment
Sample size(s) [*textboxes*]	Enter the intention to treat sample size in the corresponding textboxes: a. Total sample size b. Sample size of comparison / control group c. Sample size of intervention / treatment group If there is no control group or comparison condition, enter 'N/A' in the 'Comparison' text box.

Sample characteristics [*dropdown menu and textboxes*]	The nature of sample characteristics will depend on the units of analysis used for the evaluation. Enter sample characteristics for the overall sample.

Individuals or groups of individuals (e.g., gangs)	1. Sample age: enter in textbox provided.
	
	2. Sample gender:
	
	    a. Male
	    b. Female
	    c. Mixed
	    d. Not reported
	
	3. Sample socioeconomic status:
	
	    a. Low
	    b. Average
	    c. High
	    d. Mixed
	    e. Other (specify in adjacent textbox)
	    f. Not reported

Geographical areas	Select the size of the geographic unit from the list below:
	
	    a. Micro places (e.g., hotspots, specific addresses or street segments)
	    b. Small police-defined areas (e.g., one police station or beat)
	    c. Larger police defined areas (e.g., entire districts, sectors or states)
	    d. Neighbourhood or community
	    e. City or town
	    f. State
	    g. Other (specify in adjacent text box)

Attrition [*dropdown menu with textbox*]	Indicate whether attrition was an issue in the study by selecting an option from the list below:
	
	    a. No
	    b. Unclear
	    c. Yes
	      → Describe in textbox (e.g., how many cases were lost, how they were lost and which groups they were lost from)

Study Quality Tab

This tab asks for information about the methodological rigour of the evaluation and is used to assess potential bias of eligible studies.

Coding Field	Information

Implementation success
 	
[*dropdown menu and textbox*]
	1. Select one option from the list provided:
	   a. Intervention implemented as planned
	   b. Intervention implemented nearly as planned
	   c. Intervention not implemented or implemented in a radically different way than originally planned
	   d. Unclear or no process evaluation provided in the document
	
	2. If you select (b), briefly describe the implementation issues in the textbox provided (e.g., intervention not implemented consistently, staff turnover, lack of cooperation from third parties)

IDCG Risk of Bias Criteria [*dropdown menus*]	Use the IDCG Risk of Bias Tool to select 'Yes', 'No' or 'Unclear' for the following coding fields: 1. Mechanism of assignment 2. Group equivalence 3. Hawthorne and John Henry effects 4. Spill-overs 5. Selective outcome reporting 6. Other 7. Confidence intervals

Outcomes Tab

Because each study can have multiple outcomes, you will need to code each outcome separately. To add another outcome to a study record, click the 'Add another outcome' button and another outcome tab will appear. Only record information for outcomes that are evaluated.

Outcome Overview

Coding Field	Information
Outcome definition [*textbox*]	Using the textbox provided and the same terminology as the document, describe the outcome being measured (e.g., narcotics arrests, calls-for-service, social disorder).

Data source
 	
[*dropdown menu and textbox*]
	Specify the how the outcome data was collected by selecting one option from the list below:
	
	   a. Official data
	   b. Interview
	   c. Self-report (e.g., survey)
	   d. Systematic observation
	   e. Other (specify in adjacent text box)

Time of measurement(s)
 	
[*textbox*]
	Using the textbox provided, describe when the data was collected for this outcome. Ensure you describe the following information:
	
	   a. Number of measurements
	   b. When the measurements occurred (e.g., before or after the intervention, length of time between measurements , time between intervention and measurement)

Level of analysis
 	
[*dropdown menu*]
	Specify the outcome's level of analysis by selecting one option from the list below:
	
	   a. Individual
	   b. Groups of individuals (e.g., gangs)
	   c. Micro places (e.g., hotspots, specific addresses or street segments)
	   d. Small police-defined areas (e.g., one police station or beat)
	   e. Larger police defined areas (e.g., entire districts, sectors or states)
	   f. Neighbourhood or community
	   g. City or town
	   h. State
	   i. Other (specify in adjacent text box)

Raw difference favours
 	
[*dropdown menu*]
	Ignoring statistical significance, which condition does the raw effect/difference favour? Select one option from the list below:
	
	   a. Comparison condition
	   b. Experimental condition
	   c. Neither condition (difference = 0)
	   d. Cannot tell

Statistically significant results
 	
[*dropdown menu*]
	Indicate whether there were statistically significant differences or effects for this outcome by selecting an option from the list below:
	
	   a. Yes
	   b. No
	   c. Not tested
	   d. Cannot tell

Outcome Effect Size

Coding Field	Information
Effect size page number [*textbox*]	Specify the page number where the effect size is reported for this outcome.

How was the effect size obtained?
 	
[*dropdown menu*]
	Using the drop down menu, specify how the effect size was obtained for this outcome:
	
	   a. Reported in document
	   b. Calculated by user
	
	(If the effect size needs to be calculated but does not use the common types of data shown in the effect size calculation tabs, use the link to Wilson's online effect size calculator and calculate the effect size from the

	given data, then enter the calculated effect size data in the next step.)

Effect size data
 	
[*textboxes*]
	Identify the type of effect size reported for this outcome and then enter the required data for that effect size in the text boxes provided:
	
	1. Cohen's standardised mean difference (*d*)
	   a. d coefficient
	   b. Lower 95% CI
	   c. Upper 95%
	   d. v
	
	2. Hedge's standardised mean difference (*g*)
	   a. g coefficient
	   b. Lower 95% CI
	   c. Upper 95%
	   d. v
	
	3. Correlation coefficient (*r*)
	   a.
	      • *r* coefficient
	      • Lower 95% CI
	      • Upper 95% CI
	      • v
	      <u>OR</u>
	   b.
	      • Fisher's Z coefficient
	      • Lower 95% CI
	      • Upper 95% CI
	      • v
	
	4. Odds ratios
	   a.
	      • OR coefficient
	      • Lower 95% CI
	      • Upper 95% CI
	      <u>OR</u>
	   b.
	      • LOR coefficient
	      • Var(LOR)
	
	5. Risk ratios
	   a.
	      • RR coefficient
	      • Lower 95% CI
	      • Upper 95% CI
	      <u>OR</u>
	   b.
	      • LRR coefficient
	      • Var(LRR)

Data for Calculation of Effect Size

If there is no standardised effect size calculated for this outcome, you will need to enter the appropriate data in the 'Data for effect size calculation' tabs. The data you will need to enter will depend on the type of data that is reported in the document.

Coding Field	Data Required
Location of data [*textboxes*]	Using the textbox provided, enter the page number of the document where the data is provided (top left corner of the 'Data for effect size calculations 1' tab).

Treatment and comparison group sample sizes
 	
[*textboxes*]
	Record the following sample sizes for this effect size in the corresponding textboxes:
	
	1. Total *N*
	2. Treatment *n*
	   a. Total
	   b. Pre
	   c. Post
	3. Control *n* (enter ' - ' if there is no control group)
	   a. Total
	   b. Pre
	   c. Post
	Remember, the sample size may not be the same across different outcomes in the same study or may be different to the originally recruited sample.

Mean differences
 	
[*textboxes*]
	Record the required data in the corresponding textboxes. Where there is no data provided, enter 'N/A'.
	
	1. Treatment
	   a. Mean
	   b. SD
	   c. Mean gain
	   d. SD gain
	   e. Pre/Post *r*
	   f. Pre SD
	   g. Post SD
	   h. Paired *t*
	
	2. Control
	   a. Mean
	   b. SD
	   c. Mean gain
	   d. SD gain
	   e. Pre/Post *r*
	   f. Pre SD
	   g. Post SD
	   h. Paired *t*
	3. Full sample standard deviation

Frequencies or proportions
 	
[*textboxes*]
	Record the required data in the corresponding textboxes:
	
	1. Frequency
	   a. Treatment *n* with successful outcome ('Yes' textbox)
	   b. Treatment *n* without successful outcome ('No' textbox)
	   c. Control *n* with successful outcome ('Yes' textbox)
	   d. Control *n* without successful outcome ('No' textbox)
	
	2. Proportion
	   a. Treatment *n* with successful outcome ('Yes' textbox)
	   b. Treatment *n* without successful outcome ('No' textbox)
	   c. Control *n* with successful outcome ('Yes' textbox)
	   d. Control *n* without successful outcome ('No' textbox)

Correlation or regression coefficients
 	
[*textboxes*]
	Record the required data in the relevant corresponding textboxes:
	
	1. Unadjusted correlation
	   a. Value for coefficient *r*
	   b. *p*-value for *r*
	
	2. Standardised regression *b*
	   a. Value for coefficient *b*
	   b. *p*-value for *b*
	   c. SE *b*
	
	3. Standardised regression β
	   a. Value for coefficient β
	   b. *p*-value for β
	   c. SE of β
	
	4. Standard deviation of outcome
	
	5. Control variables
	
	6. Number of control variables
	
	7. Point biserial correlation
	   a. Value for point biserial correlation *r*
	   b. *p*-value for point biserial *r*
	
	8. Phi coefficient
	   a. Value for Phi coefficient
	   b. *p*-value for Phi

Statistical significance tests
 	
[*Textboxes*]
	Record the required data in the relevant corresponding textboxes:
	
	1. *t* statistic
	   a. *t*-value
	   b. *p*-value for *t*
	
	2. *F* statistic
	   a. *F*-value
	   b. *p*-value for *F*

	3. Chi-square statistic
	    a. Chi-square value
	    b. *p*-value for chi-square
	    c. *df* for chi-square
	
	4. Proportion of full sample with event

ANCOVA
 	
[*Textboxes*]
	Record the required data in the relevant corresponding textboxes:
	
	    a. MS-error
	    b. Correlation of covariate with DV
	    c. Treatment mean
	    d. Treatment *n*
	    e. Control mean
	    f. Control *n*

ANOVA
 	
[*Textboxes*]
	Record the required data in the relevant corresponding textboxes:
	
	    a. F-value, treatment factor
	    b. F-value, other factor
	    c. F-value, interaction
	    d. MS-error
	    e. Treatment mean
	    f. Treatment *n*
	    g. Control mean
	    h. Control *n*

Other Information [*Textbox*]	Record any additional information that you think is relevant to calculation of the effect size for this outcome.

APPENDIX G: Adapted IDCG Risk of Bias Tool

Tool to assess risk of bias and internal validity of social experiments and quasi-experiments[7]

The following tool enables the consistent assessment of internal validity of social experiments and quasi-experiments including randomised control trials (RCTs), regression discontinuity designs (RDDs), non-randomised studies based on participant self-selection (panel data models, propensity score and covariate matching, and cross-sectional regression), and studies using instrumental variables estimation for causal identification (IV). The tool consists of eight evaluation criteria to identify threats to validity that may arise due to the following sources: selection bias, confounding, motivation bias, performance bias, outcome reporting bias, analysis reporting bias, other sources of bias, and threats to the correct calculation of statistical significance of the effect. Application of the tool is likely to require advanced knowledge of statistics and econometrics.

1. *Mechanism of assignment: was the allocation or identification mechanism able to control for selection bias?*

Type of Assignment	Score 'YES' if…	Score 'UNCLEAR' if…	Score 'NO' if…
Randomised (e.g., RCT)	• A random component in the sequence generation process is described (e.g., referring to a random number table)[8]; • And if the unit of allocation was at group level (geographical/ social/ institutional unit) and allocation was performed on all units at the start of the study, • Or if the unit of allocation was by beneficiary or group and there was some form of centralised allocation mechanism such as an on-site computer system;	• The document does not provide details on the randomisation process, or uses a quasi-randomisation process for which it is not clear has generated allocations equivalent to true randomisation.	• Any failure in the allocation mechanism could affect the randomisation process[9].

---

[7] The tool has been adapted from an instrument developed by Jorge Hombrados and Hugh Waddington, drawing on existing tools, in particular EPOC (n.d.), Higgins and Green (2011) and Coalition for Evidence-Based Policy (2010). Thanks to Richard Palmer-Jones, Maren Duvendack and Phil Davies for comments on previous drafts.

[8] If a quasi-randomized assignment approach is used (e.g. alphabetical order), you must be sure that the process truly generates groupings equivalent to random assignment, to score "Yes" on this criteria. In order to assess the validity of the quasi-randomization process, the most important aspect is whether the assignment process might generate a correlation between participation status and other factors (e.g., gender, socioeconomic status) determining outcomes; you may consider covariate balance in determining this (see question 2).

[9] If there are serious concerns about the randomisation process or the group equivalence, assess the risk of bias of the study using the relevant questions for the appropriate methods of analysis (cross-sectional regressions, difference-in-difference, etc) rather than the RCTs questions.

Type of Assignment	Score 'YES' if…	Score 'UNCLEAR' if…	Score 'NO' if…
Discontinuity (e.g., Regression discontinuity designs)	• Allocation is made based on a pre-determined discontinuity on a continuous variable (regression discontinuity design) and blinded to participants; or • If not blinded, individuals reasonably cannot affect the assignment variable in response to knowledge of the participation decision rule;	• The assignment variable is either non-blinded or it is unclear whether participants can affect it in response to knowledge of the allocation mechanism.	• There is evidence that participants altered the assignment variable prior to assignment[10].
Non-random or self-selection (e.g., matching, regression analysis – excluding IV)	• Participants and non-participants are either matched based on all relevant characteristics explaining participation and outcomes, or • All relevant characteristics are accounted for.[11] [12]	• It is not clear whether all relevant characteristics (only relevant time varying characteristics in the case of panel data regressions) are controlled.	• Relevant characteristics are omitted from the analysis.
Identification based on instrumental variable (e.g., instrumental variable estimation)	• An appropriate instrumental variable is used which is exogenously generated: e.g. due to a 'natural' experiment or random allocation.	• The exogeneity of the instrument is unclear (both externally as well as why the variable should not enter by itself in the outcome equation).	Score 'No' otherwise.

---

[10] If there are serious concerns with the assignment process or the group equivalence, to assess the risk of bias of the study using the relevant questions for the appropriate methods of analysis (cross-sectional regressions, difference-in-difference, etc) rather than the RDDs questions.

[11] Accounting for and matching on all relevant characteristics is usually only feasible when the programme allocation rule is known and there are no errors of targeting. It is unlikely that studies not based on randomisation or regression discontinuity can score "YES" on this criterion.

[12] There are different ways in which covariates can be taken into account. Differences across groups in observable characteristics can be taken into account as covariates in the framework of a regression analysis or can be assessed by testing equality of means between groups. Differences in unobservable characteristics can be taken into account through the use of instrumental variables (see also question 1.d) or proxy variables in the framework of a regression analysis, or using a fixed effects or difference-in-differences model if the only characteristics which are unobserved are time-invariant.

2. *Group equivalence: was the method of analysis executed adequately to ensure comparability of groups throughout the study and prevent confounding?*

Type of Design	Score 'YES' if…	Score 'UNCLEAR' if…	Score 'NO' if…
RCTs or quasi-RCTs[13]	• Baseline characteristics of the study and control/comparisons are reported and overall[14] similar based on *t*-test or ANOVA for equality of means across groups, • or covariate differences are controlled using multivariate analysis; • And the attrition rates (losses to follow up) equivalent across treatment and control, or the study assesses that loss to follow up units are random draws from the sample (e.g., by examining correlation with determinants of outcomes, in both treatment and comparison groups); • And problems with cross-overs and drop outs are dealt with using intention-to-treat analysis or in the case of drop outs, by assessing whether the drop outs are random draws from the population; • And, for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme (e.g., weather, infrastructure, community fixed effects, etc) through multivariate analysis.	• Insufficient details are provided on covariate differences or methods of adjustment; • Or insufficient details are provided on cluster controls.	Score 'No' otherwise
Regression Discontinuity	• The interval for selection of treatment and control group is reasonably small, • Or authors have weighted the matches on their distance to the cut-off point, • And the mean of the covariates of the individuals immediately at both sides of the cut-off point (selected sample of participants and non-participants) are overall not statistically different based on t-test or ANOVA for equality of means, • Or significant differences have been controlled in multivariate analysis; • And, for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme (e.g., weather, infrastructure, community fixed effects, etc) through multivariate analysis.	• There are covariate differences across individuals at both sides of the discontinuity which have not been controlled for using multivariate analysis, or if insufficient details are provided on controls, • Or if insufficient details are provided on cluster controls.	Score 'No' otherwise

---

[13] Please note that when a), b), or f) score no or large differences in baseline characteristics, we suggest assessing risk of bias considering other study design (Diff-in-Diff, cross-sectional regression, instrumental variables)

[14] Even in the context of RCTs, when randomisation is successful and carried out over sufficiently large assignment units, it is possible that small differences between groups remain for some covariates. In these cases, study authors should use appropriate multivariate methods to correcting for these differences.

Type of Design	Score 'YES' if…	Score 'UNCLEAR' if…	Score 'NO' if…
RCT using difference-in-difference methods of analysis	• The authors use a difference-in-differences (or fixed effects) multivariate estimation method; • The authors control for a comprehensive set of time-varying characteristics;[15] • And the attrition rate is similar in treatment and control, or the study assesses that drop-outs are random draws from the sample (e.g., by examining correlation with determinants of outcomes, in both treatment and comparison groups); • And, for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme (e.g., weather, infrastructure, community fixed effects, etc) through multivariate analysis.	• Insufficient details are provided, • Or if insufficient details are provided on cluster controls.	Score 'No' otherwise
Statistical matching studies (e.g., PSM and covariate matching)	• Matching is either on baseline characteristics or time-invariant characteristics which cannot be affected by participation in the programme; and the variables used to match are relevant (e.g. demographic and socio-economic factors) to explain both participation and the outcome (so that there can be no evident differences across groups in variables that might explain outcomes) (see fn. 6). • In addition, for PSM Rosenbaum's test suggests the results are not sensitive to the existence of hidden bias. • And, with the exception of Kernel matching, the means of the individual covariates are equated for treatment and comparison groups after matching; • And, for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme (e.g., weather, infrastructure, community fixed effects, etc) through multivariate or any appropriate analysis.	• Relevant variables are not included in the matching equation, or if matching is based on characteristics collected at endline, • Or if insufficient details are provided on cluster controls.	Score 'No' otherwise
Regression-based studies using cross-section data (excluding IV)	• The study controls for relevant confounders that may be correlated with both participation and explain outcomes (e.g. demographic and socio-economic factors at individual and community level) using multivariate methods with appropriate proxies for unobservable covariates (see fn. 13), • And a Hausman test[16] with an appropriate instrument suggests there is no evidence	• Relevant confounders are controlled but appropriate proxy variables or statistical tests are not reported,	Score 'No' otherwise

---

[15] Knowing allocation rules for the programme – or even whether the non-participants were individuals that refused to participate in the programme, as opposed to individuals that were not given the opportunity to participate in the programme – can help in the assessment of whether the covariates accounted for in the regression capture all the relevant characteristics that explain differences between treatment and comparison.

[16] The Hausman test explores endogeneity in the framework of regression by comparing whether the OLS and the IV approaches yield significantly different estimations. However, it plays a different role in the different methods of analysis. While in the OLS regression framework the Hausman test mainly explores endogeneity and therefore is

Type of Design	Score 'YES' if…	Score 'UNCLEAR' if…	Score 'NO' if…
	of endogeneity, • And none of the covariate controls can be affected by participation; • And either, only those observations in the region of common support for participants and non-participants in terms of covariates are used, or the distributions of covariates are balanced for the entire sample population across groups; • And, for cluster-assignment, authors control particularly for external cluster-level factors that might confound the impact of the programme (e.g., weather, infrastructure, community fixed effects, etc) through multivariate analysis.	• Or if insufficient details are provided on cluster controls.	
Instrumental variable approaches	• The instrumenting equation is significant at the level of $F \geq 10$ (or if an F test is not reported, the authors report and assess whether the R-squared (goodness of fit) of the participation equation is sufficient for appropriate identification); • The identifying instruments are individually significant ($p \leq 0.01$); for Heckman models, the identifiers are reported and significant ($p \leq 0.05$); • Where at least two instruments are used, the authors report on an over-identifying test ($p \leq 0.05$ is required to reject the null hypothesis); and none of the covariate controls can be affected by participation and the study convincingly assesses qualitatively why the instrument only affects the outcome via participation[17]. • And, for cluster-assignment, authors particularly control for external cluster-level factors that might confound the impact of the programme (e.g., weather, infrastructure, community fixed effects, etc) through multivariate analysis.	• Relevant confounders are controlled but appropriate statistical tests are not reported or exogeneity[18] of the instrument is not convincing, • Or if insufficient details are provided on cluster controls (see category f) below).	Score 'No' otherwise

related with the validity of the method, in IV approaches it explores whether the author has chosen the best available strategy for addressing causal attribution (since in the absence of endogeneity OLS yields more precise estimators) and therefore is more related with analysis reporting bias.

[17] If the instrument is the random assignment of the treatment, the reviewer should also assess the quality and success of the randomisation procedure in part a).

[18] An instrument is exogenous when it only affects the outcome of interest through affecting participation in the programme. Although when more than one instrument is available, statistical tests provide guidance on exogeneity (see background document), the assessment of exogeneity should be in any case done qualitatively. Indeed, complete exogeneity of the instrument is only feasible using randomised assignment in the context of an RCT with imperfect compliance, or an instrument identified in the context of a natural experiment.

3. Hawthorne and John Henry effects: was the process of being observed causing motivation bias?

Score 'YES' if either…	Score 'UNCLEAR' if…
a) For data collected in the context of a particular intervention trial (randomised or non-randomised assignment), the authors state explicitly that the process of monitoring the intervention and outcome measurement is blinded, or argue convincingly why it is not likely that being monitored in ways that could affect the performance of participants in treatment and comparison groups in different ways. b) The study is based on data collected in the context of a survey, and not associated with a particular intervention trial, or data are collected in the context of a retrospective (ex post) evaluation.	• It is not clear whether the authors use an appropriate method to prevent Hawthorne and John Henry Effects (e.g. blinding of outcomes and, or enumerators, other methods to ensure consistent monitoring across groups).
Score 'NO' otherwise.	

4. Spill-overs: was the study adequately protected against performance bias?

Score 'YES' if…	Score 'UNCLEAR' if…	Score 'NO' if…
• The intervention is unlikely to spill-over to comparisons (e.g. participants and non-participants are geographically and/or socially separated from one another and general equilibrium effects are unlikely)[19].	• Spill-overs are not addressed clearly	• Allocation was at individual or household level and there are likely spill-overs within households and communities which are not controlled for in the analysis; • Or if allocation at cluster level and there are likely spill-overs to comparison clusters.

---

[19] Contamination, that is differential receipt of other interventions affecting outcome of interest in the control or comparison group, is potentially an important threat to the correct interpretation of study results and should be addressed via PICO and study coding.

The Campbell Collaboration | www.campbellcollaboration.org

5. Selective outcome reporting: was the study free from outcome reporting bias?

Score 'YES' if…	Score 'NO' if…
● There is no evidence that outcomes were selectively reported (e.g., all relevant outcomes in the methods section are reported in the results section).	● Some important outcomes are subsequently omitted from the results or the significant and magnitude of important outcomes was not assessed.
Score 'UNCLEAR' otherwise.	

6. Selective analysis reporting: was the study free from analysis reporting bias?

Score 'YES' if…	Score 'NO' if…
● Authors use 'common methods'[20] of estimation and the study does not suggest the existence of biased exploratory research methods[21].	● Authors use uncommon or less rigorous estimation methods such as failure to conduct multivariate analysis for outcomes equations where it is has not been established that covariates are balanced.
Other Specific Estimation Methodologies	

Research Design	Score 'YES' if…	Score 'UNCLEAR' or 'NO' if…
PSM and Covariate Matching	● Where over 10% of participants fail to be matched, sensitivity analysis is used to re-estimate results using different matching methods (Kernel Matching techniques). ● For matching with replacement, no single observation in the control group is matched with a large number of observations in the treatment group.	*Where not reported or not convincing, score "UNCLEAR". Otherwise, Score "NO".*
IV Models (incl. Heckman)	● The authors test and report the results of a Hausman test for exogeneity ($p \leq 0.05$ is required to reject the null hypothesis of exogeneity). ● The coefficient of the selectivity correction term (Rho) is significantly different from zero ($p<0.05$) (Heckman approach).	
Studies using Multivariate Regression Analysis	● Authors conduct appropriate specification tests (e.g., reporting results of multicollinearity test, testing robustness of results to the inclusion of additional variables, etc).	

---

[20] 'Common methods' refers to the use of the most credible method of analysis to address attribution given the data available.
[21] A comprehensive assessment of the existence of 'data mining' is not feasible particularly in quasi-experimental designs where most studies do not have protocols and replication seems the only possible mechanism to examine rigorously the existence of data mining.

7. *Other: was the study free from other sources of bias?*

Important additional sources of bias may include: concerns about blinding of outcome assessors or data analysts; concerns about blinding of beneficiaries so that expectations, rather than the intervention mechanisms, are driving results (detection bias or placebo effects)[22]; concerns about courtesy bias from outcomes collected through self-reporting; concerns about coherence of results; data on the baseline collected retrospectively; information is collected using an inappropriate instrument (or a different instrument/at different time/after different follow up period in the comparison and treatment groups).

Score 'YES' if…	Score 'UNCLEAR' if…	Score 'NO' if…
• The reported results do not suggest any other sources of bias.	• Other important threats to validity may be present	• It is clear that threats to validity are present and not controlled for.

8. *Confidence intervals*

NOTE: for full internal validity assessment – i.e., risk of bias in effects and precision based on true confidence intervals (Type I error, Type II error) – assessment should include the following:

Type of Study	Score 'YES' if…	Score 'UNCLEAR' if…	Score 'NO' if…
Studies using parametric regression models, such as OLS (distribution of error term and heteroscedasticity)	• The authors test and fail to reject the null of homoscedasticity (e.g. through a Breusch-Pagan test for heteroscedasticity ($p>0.05$)) and test for the assumed error distribution (e.g. Kolmogorov-Smirnov test for non-normality ($p>0.05$)) • Or if the test suggests the existence of heterogeneity or non-normality, the study corrects for them (e.g. use of log transformation in the dependent variable).	• Results of any test are not reported.	Score 'No' otherwise[23]
If, despite large effects, the study fails to find the	• The sample size is enough to detect a relevant significant	• It is not clear	• The same is not

---

[22] All interventions may create expectations (placebo effects), which might confound causal mechanisms. In social interventions, which usually require behaviour change from participants, expectations may form an important component of the intervention, so that isolating expectation effects from other mechanisms may be less relevant.

[23] Standard errors may be inflated in parametric approaches if the intervention does not have a homogeneous effect across the whole sample population, and the authors fail to conduct appropriate sub-group analyses.

Type of Study	Score 'YES' if…	Score 'UNCLEAR' if…	Score 'NO' if…
effects significant (power of the study)	effect	whether the sample size is sufficiently large enough to detect medium or large significant effects	sufficiently large enough to detect medium or large significant effects.
Clustered studies (unit of analysis error)	• The analysis is carried out at the relevant unit of treatment assignment, • Or the study accounts for lack of independence between observations within assignment clusters.	• The study does not report enough information on the unit of treatment assignment	• The analysis is carried out at different unit than the assignment.