Journal of Criminal Justice 92 (2024) 102195

Contents lists available at ScienceDirect

Journal of Criminal Justice

journal homepage: www.elsevier.com/locate/jcrimjus

A theory of change driven approach to evaluating a multi-agency stalking intervention programme

Jyoti Belur [a,*], Lisa Tompson [b], Kritika Jerath [c]

[a] Department of Security and Crime Science, University College London, 35 Tavistock Square, London WC1H 9EZ, UK
[b] Institute of Security and Crime Science, University of Waikato, Hamilton, New Zealand
[c] School of Sociology and Social Policy, University of Nottingham, University Park, Nottingham NG7 2RD, UK

ARTICLE INFO

Keywords:
Multi-agency partnership
Evaluation
Stalking
Theory of change
EMMIE

ABSTRACT

Purpose: Multi-agency initiatives as a response to complex crimes, such as stalking, pose conceptual and operational challenges for practitioners and evaluators. This study addresses these challenges, by combining a theory of change driven approach with the realist-inspired EMMIE evaluation framework to present findings from a pilot multi agency stalking intervention partnership in England and Wales.
Methods: The study uses a mixed methods approach based on analysis of police data and semi-structured interviews with stakeholders to present a process and effect evaluation of a complex social intervention.
Results: Findings based on a small sample and limited follow up period show nonsignificant results in terms of reducing the harmful effects of stalking, but which are encouragingly in the right direction. The case study demonstrates the importance of a theory driven approach to multi-agency evaluation and identifies essential factors necessary for ensuring implementation success.
Conclusions: A theory of change driven approach and an EMMIE-inspired evaluation can help identify whether interventions work, how they work, for whom they work, and under what conditions. It demonstrates the importance of data collection and provision for long term evaluation plans, especially for complex social interventions. The paper makes a significant methodological and empirical contribution to evaluation literature.

1. Introduction

Despite growing public interest in multi-agency partnerships and their potential to address complex problems, such as crime-related ones (Cox, 2019; Vander Laenen & Persac, 2014), evaluation efforts to support the seemingly universal belief in the effectiveness of multi-agency approaches are underdeveloped (Chapman, Higgins, & Hales, 2017; Mazerolle, Cherney, Eggins, Hine, & Higginson, 2021). Most initiatives in this area assume that a multi-agency approach is a more effective, holistic, response to crime, as opposed to reliance solely on law enforcement agencies. However, such assumptions are rarely tested. Instead, findings from cogent fields indicate that the effectiveness of interagency collaboration is mixed (Morgan, Pullon, Garrett, & McKinlay, 2019). Further, Gray (2014) suggests that whilst multi-agency partnerships might have good intentions, they are often beset by challenges, internal tensions, and lack quality assessment of services provided, to be as successful as expected.

Some studies suggest that multi-agency work is no more effective than stand-alone individual agencies (Atkinson, Jones, & Lamont, 2007; Cleaver, Maras, Oram, & McCallum, 2019; Herbert & Bromfield, 2019). However, they caution that the diversity in definitions, approaches, and contexts, as well as the different levels of multi-agency collaboration (Horwath & Morrison, 2007), must be explored further to identify mechanisms associated with effective implementation. This study fills a gap in the evaluation literature on the effectiveness of multi-agency partnerships by presenting a theory-driven evaluation based on the EMMIE framework (Johnson, Tilley, & Bowers, 2015) of one police force area involved in a pilot project titled 'Multi-agency Stalking Intervention Programme' (MASIP) in England and Wales (E&W) conducted between 2018 and 2020.

2. Stalking: A complex social problem

Stalking is a complex and high-harm crime, with heterogenous aetiologies (MacKenzie & James, 2011; Nijdam-Jones, Rosenfeld, Gerbrandij, Quick, & Galietta, 2018). Estimates of prevalence rates of

* Corresponding author.
E-mail addresses: j.belur@ucl.ac.uk (J. Belur), ltompson@waikato.ac.nz (L. Tompson), Kritika.Jerath@nottingham.ac.uk (K. Jerath).

https://doi.org/10.1016/j.jcrimjus.2024.102195
Received 13 February 2024; Received in revised form 13 May 2024; Accepted 13 May 2024
Available online 24 May 2024

offline and online stalking victimization vary but are generally held to be widespread (Björklund, Häkkänen-Nyholm, Sheridan, & Roberts, 2010; McNamara & Marsil, 2012; Owens, 2016). Whilst no universal definition of stalking yet exists, it is commonly conceptualized as 'repeated, unwanted, contact that causes the recipient distress and/or fear' (Korkodeilou, 2016; McEwan, Harder, Brandt, & de Vogel, 2020). Thus, it is victim-defined, context specific, and consequently, difficult to identify.

Stalking manifests in varied and complex ways, making it a difficult crime to respond to (Storey & Hart, 2011). Much of it is hidden from official statistics, as a large proportion of stalking incidents go unreported (Jerath, Tompson & Belur, 2022, 2023). Victims who do report often have negative experiences because the police lack a nuanced understanding of stalking behaviour (Holt, Lee, Liggett, Holt, & Bossler, 2019). The severity of stalking harms is similarly underappreciated by wider society (McKeon, McEwan, & Luebbers, 2015).

Although legal sanctions might intuitively seem the most effective way to deal with stalking behaviour, clinicians who treat perpetrators claim that these alone do not address their fixation and obsession (Mullen, Pathé, & Purcell, 2000). Other research has indicated that the unwanted contact can escalate, rather than be diminished, after legal sanctions (Ostermeyer, Friedman, Sorrentino, & Booth, 2016). Indeed, guidance in E&W[1] no longer advises police cautions to be used as a sanction in stalking cases; instead, the UK Stalking Protection Act (2019) introduced protection orders as an early intervention mechanism for the police to put protective measures in place for victims who report stalking incidents (Kelly, 2020).

It is well documented that people who stalk have complex needs (HMICFRS and HMCPSI, 2017; McEwan, Shea, Nazarewicz, & Senkans, 2017) and clinical interventions form the nascent evidence base into preventing stalking recidivism to date (Wheatley, 2019). However, even after clinical interventions have been delivered, reoffending rates remain stubbornly high at 33–50% (MacKenzie et al., 2009; McEwan et al., 2017). Measurement of stalking reoffending poses greater challenges than other types of offending (Walby, 2005) due to its protracted nature and because subsequent reporting depends on victim satisfaction with the criminal justice response following their initial reporting. Since stalking is a highly nuanced social problem which poses reoffending risks, multi-agency partnerships (hereafter partnerships) appear to be a viable way of bringing together requisite expertise to effectively manage the risks and harms of stalking (Pathé et al., 2018; Rosenbaum, 2002).

3. Conceptual underpinnings of multi-agency working

Several assumptions underlie conventional wisdom across social and health policy areas that multi-agency working is superior to single agency working. Partnerships are believed to create efficiency gains across the public sector by avoiding duplication of efforts across different agencies (Sloper, 2004). In addition, and *crucially,* they are also believed to work more effectively in enhancing service provision than agencies operating alone (Hester, Eisenstadt, Jones, & Morgan, 2017). Reasons given are expedited information sharing, improved referral processes, integration of working practices across organisational boundaries, access to professional expertise, and the ability to present a unified response to service users (Berry, Briggs, Erol, & van Staden, 2011; Burnett & Appleton, 2004; Robinson, 2004; Sloper, 2004). Partnerships that focus on managing high risk people can potentially provide a better *quality* of response to service users – due to the synthesis of diverse perspectives on how to respond with the potential to target multiple causal mechanisms driving the problem – and/or a better *quantity* (i.e., dosage) of response to service users (Hester et al., 2017; Robinson & Clancy, 2021; Rosenbaum, 2002). For partnerships responsible for managing high-risk offenders or situations, the gains in effectiveness can be the difference between a calamitous outcome and a less harmful one. Echoing others in the risk management field, Reeves (2013: 41) asserts that there has been a "growing realization that individual risks cannot be managed by focused work undertaken by single agencies working independently to their own aims".

4. Role of multi-agency partnerships in managing risk and reducing crime

There are several putative benefits of crime reduction partnerships, and practitioners value them highly (Berry et al., 2011). Furthermore, there are many claims in the literature that multi-agency working delivers greater support and benefits for victims (Lea and Callaghan, 2016), which is a laudable outcome. However, these benefits do not inevitably translate to crime reduction (Chapman et al., 2017). Crime reduction partnerships do not exist in a vacuum. They are embedded within the intricate and interrelated social systems that they are attempting to change. Evaluations of partnerships are consequently thwarted by the complexity of the social initiatives they seek to investigate.

It is difficult to evaluate crime reduction partnerships because their operations are dynamic and responsive. Different intervention decisions are taken based on dynamic risk assessments, as risk varies. Partnerships can tailor bespoke *combinations of interventions* to individual cases or problems, depending on the nature of said case. Consequently, it becomes challenging to isolate individual treatment effects and to divorce them from the multifaceted social contexts in which they exist. This lack of optimal conditions for traditional experiments is likely the reason for the lack of rigorously designed evaluations of multi-agency partnerships (Rosenbaum, 2002). We contend that an alternate evaluation approach would be instrumental in advancing this evidence base, particularly one that has a theory of change at its core.

5. Theory of change and programme success

A theory of change (ToC) is 'a theory of how and why an initiative works' (Weiss, 1997) which 'can be empirically tested by measuring indicators for every expected step on the hypothesized causal pathway to impact' (De Silva et al., 2014: 4). The term has emerged from evaluation settings characterised by 'uncertainty and emergence' (Van Tulder & Keen, 2018). Scholars developed Weiss (1997) initial definition further to suggest that it needs to provide a framework that explicates the links between activities, outcomes, and the context of the initiative (Connell & Kubisch, 1998), along with the assumptions and risks which may affect the activities (Barnes, Matka, & Sullivan, 2003).

For a ToC to be considered sufficient, it must be plausible, feasible, and testable (Connell & Kubisch, 1998). The plausibility of a ToC is concerned with the logical sense which ties a chain of events leading to an outcome. Feasibility is most important when considering resources, and support, against a political/institutional backdrop which will make an initiative possible to carry out. A ToC driven approach to evaluation provides a promising framework for evaluating complex social interventions, as it helps clarify what should be evaluated, and strengthens the credibility of its conclusions (Connell & Kubisch, 1998).

A ToC is typically developed by evaluators, in collaboration with stakeholders, and adjusted throughout the implementation of an initiative and evaluation via an ongoing awareness of and reflection on exploring change and explaining how it happens (Connell & Kubisch, 1998). ToC models are usually characterised as a logic model which links activities to outcomes and tries to capture causation of change. Without acknowledging the causal mechanisms involved in complex change processes, it is difficult to identify the contributors to a successful outcome (Dyson & Todd, 2010) and therefore different aspects affecting

[1] The UK Crown Prosecution Service's advice has, since 2009, increasingly erred towards encouraging cautions to be used sparingly. Cautioning stalkers is inappropriate for several reasons, least not that they deny the victim recourse to a restraining order.

change must be analysed. Blamey and Mackenzie (2007) further note that a ToC should inform an evaluation's purpose, key questions, and consequently, the selection of research methods.

A ToC approach fits neatly within the realist evaluation movement (Pawson & Tilley, 1997). This approach argues that context-mechanism-outcome patterns are observable, and that these generate insight about the conditions under which initiatives might be effective. Whilst realist evaluation has become common in health and social policy evaluation spheres, it has yet to be applied to studying the impacts of multi-agency partnerships. However, it dovetails with Rosenbaum (2002) suggestion for an optimal evaluation of a crime reduction partnership.

The present study attempts to address some of the shortcomings identified in the evaluation literature of previous multi-agency partnerships by adopting a ToC driven approach (Connell & Kubisch, 1998) and using the EMMIE evaluation framework (Johnson et al., 2015), which is inspired by 'realist evaluation' principles. The EMMIE framework is concerned with drawing out not just whether an initiative 'works' but how, for whom, and under what conditions. This framework was designed to capture information (and interactions between) the **E**ffect of an initiative, the **M**echanisms causing the effect, the **M**oderating (i.e., contextual) conditions that activate the mechanisms, crucial information on **I**mplementation and, finally, data on the **E**conomic cost and benefits of delivering the initiative.

We applied this evaluation framework to the Multi-agency Stalking Intervention Programme (MASIP), a well-resourced pilot project set up in 2018 to reduce the risks and harms associated with stalking across three police areas. In this paper we focus on presenting findings from one pilot site that collected data robust enough to enable an attempt at assessing whether the intervention had reduced reoffending.[2] This pilot site operated over a subset of the geographical area covered by Cheshire Police, UK and was called the Integrated Anti-Stalking Unit (IASU). We seek to answer the following research questions:

1. What was the IASU's ToC and through what mechanisms were the interventions intended to work?
2. Was the IASU effective at reducing reoffending?
3. What implementation factors were essential for IASU's ToC?

Consequently, in this paper we mainly report on three aspects of the EMMIE framework – Effect, Mechanisms, and Implementation.[3]

6. Methods

6.1. The study setting

The MASIP was a proof-of-concept initiative which aimed to reduce the impact of stalking on victims by developing a partnership intervention model. The initiative sought to draw expertise and knowledge across the multi-agency spectrum to inform the risk management process associated with managing stalking cases. Key partners included police, probation, health, and victim advocates.

Risk management plans were devised for both victims and perpetrators, with victims receiving support from victim advocates throughout the entire case management process. A subset of perpetrators was offered therapeutic interventions to address their obsessive and fixated behaviours when, a) they were not already engaged with other agencies (e.g., health), and b) treatment was deemed to be clinically indicated by health professionals. A ToC approach to evaluation using the EMMIE evaluation framework demanded a mixed methods approach as it combined a process evaluation (understanding how the intervention was set up, developed, changed, and what activities were undertaken), with an outcome evaluation (exploring whether the intervention works). Hence, we employed a variety of data collection methods and analytic strategies.

6.2. Qualitative analysis

Semi-structured interviews were conducted with six of the core IASU partnership staff and two members of the central project management team at the beginning and end of the initiative. The first wave of interviews explored the aims, barriers, and facilitators for the project, and identified the conceptual elements of the ToC model. The second wave of interviews conducted months later, probed the experiential learning from stakeholders on the benefits and challenges multi-agency working brought to the case management.

Interviews lasted between 10 and 75 minutes and were recorded with the participant's permission. The interviews were anonymized, professionally transcribed, and were coded and thematically analysed using qualitative software NVIVO. Thematic analysis focused on pre-identified themes that answered the research questions and emerging themes were incorporated as the analysis progressed. The data was coded by one member and checked by another member of the evaluation team and analysed jointly. Conclusions were drawn from discussion between the evaluation team members and were refined by feedback from the project stakeholders.

The evaluation team also observed meetings, clinics, and triage processes in the pilot site and held informal discussions with stakeholders to clarify understanding of the processes involved. Documentation of the operating model, provided by the unit, was reviewed. Data collected through various methods were triangulated to deepen reliability - for example, interview data was supported or qualified through our observations and analysis of documents.

The combination of qualitative methods used over the fieldwork informed the generation of a priori hypotheses (see Jamal et al., 2015). Hypotheses related to the *mechanisms* (the way in which IASU exerted its effects), the *moderators* (the pre-existing contextual conditions) and the *implementation* (conditions introduced by IASU and the evaluation). Hypotheses were generated before quantitative data were collected, to protect against subjectivity.

6.3. Quantitative analysis

Comprehensive quantitative partnership data was provided by IASU. This contained information on the perpetrator and victim, details of the case, details of what interventions were used during the case management, and a field that stated whether further stalking behaviour had been reported at three-monthly intervals. The stalking behaviour could be reported by victims to the police or to their victim advocate, by probation, or could come to the unit's attention via other means. Consequently, it is the most comprehensive measure of stalking reoffending of which we are aware in the literature, made possible by the cohesive multi-agency partnership.

In line with clinical guidance, we defined reoffending as additional stalking behaviour within six months of the original offence coming to the attention of the IASU (McEwan, Mullen, & MacKenzie, 2009). However, due to the severely restricted evaluation timescales (17-months with no follow-up period funded) we deviate from research that uses the conclusion of a health intervention as the starting point for measurement. This means that we are liable to overestimate reoffending since we start measuring at a much earlier point (the antecedent

[2] The other two pilot sites struggled with the data recording and sharing systems across partner agencies in their respective areas and are therefore not being presented in this paper (see Tompson, Belur, & Jerath, 2020). This means that the unit represented in this paper is not necessarily representative of other multi-agency groups set up to address stalking. Indeed, this police area had working processes that had matured over time prior to MASIP beginning. The results presented herein should therefore not be taken as generalisable to other multi-agency groups.

[3] There was scant data to test some of the moderators and findings on costs are presented elsewhere (Tompson, Belur, & Jerath, 2021)

offence), regardless of what stage the intervention was at the time the analysis was conducted. Secondary criminal justice outcome data on stalking offences was secured from a Police force judged to be most like Cheshire.[4] Due to the timing of the data request, this was for the period of August 2016–December 2019, inclusive, and hence does not include the full study period.

Descriptive analysis was undertaken of partnership data and the secondary criminal justice outcome data, including chi-square and Fisher's exact tests. Logistic regression was run with the reoffending outcome as the dependent variable and intervention types by the IASU as independent variables. Due to the modest sample size ($n = 71$), we ran two versions of the logistic regression model; one standard and one with Firth (1993) method, that acts in a similar way to penalized likelihood to correct for biases associated with small samples. Prior to running the models, we first dichotomized the independent variables so that they were mutually exclusive.

7. Findings

Cases were referred into the IASU by a range of agencies and were initially screened and triaged. Regular internal meetings to discuss cases were facilitated by the co-location of core staff, and these supported the functioning of the fortnightly partnership meeting. The partnership meeting was where wider partners (e.g., Probation, Health, Social Services) were invited to risk assess specific cases, and bespoke advice was dispensed to referring agencies (e.g., investigative advice to the police). These partnership meetings were described by one interviewee as,

"In the multi-agency meeting to assess the risk, decisions are made. We share information first and then we go through our clinic assessment form methodically, so we look at the risk and then we assess that together, so the partners all put their opinion in of whether they think, is it a consensus, is it stalking, what the risks are, what the typology is? We all input. The partners don't just share information, they also input on our decision-making regarding risk and that we share that risk." (CSs03, Social Services).

Complementary tasks undertaken at these meetings included signposting to appropriate agencies and the creation of safeguarding plans for victims and perpetrators. The IASU sought to identify the most appropriate interventions to manage the risk in a case, which often integrated legal sanctions with inputs from mental health services. Victims and perpetrators were hence supported directly and indirectly within this partnership model.

We present our findings in response to the three questions guiding this research.

7.1. What were IASU's ToC and mechanisms for change?

Unusually for a multi-agency initiative (Rosenbaum, 2002), stakeholders were encouraged during the project set-up phase to develop their individual ToC by the project managers. IASU stakeholders presented their initial ToC and we used this to populate our input-output-outcome-impact model for IASU. However, based on our observations and following interviews with stakeholders, we refined their ToC to depict how the MASIP model was intended to work to achieve articulated outcomes (see Fig. 1). Our experience indicated that the partnership found it difficult to succinctly articulate the input-output-outcome-and-impact model we supplied. This is because practitioners tend to think only in terms of what actions they undertake and the results they hope to achieve. Thus, boundaries between what Rogers (2014) would traditionally call inputs, outputs, outcomes, and impact dimensions are often ambiguous. To elaborate, when asked what would be considered successful outcomes, interviewees made a clear distinction between what they considered to be successful outcomes for the unit, as well as what they would consider successes for their own agency, but did not distinguish outputs (raising awareness), instrumental outcomes (delivering training), interim outcomes (appropriate classification and charging, and increased convictions) and intended outcomes (reduced reoffending) in counting their successes. For instance, perceptions of success ranged from "generating masses of data for the public good" (CPo2, Police), to leading the perpetrator towards self-regulating their behaviour, as one interviewee said,

"For them to have worked, with our guidance, to a place where they're in employment and they haven't been in two years, they're not drinking alcohol anymore, using more appropriate coping strategies to manage emotions, being aware of their emotions and recognising when they're experiencing different things and being aware of their own triggers or what might cause them to reoffend, I think that might be considered a success". (CH14, Health/Outreach).

The notion of successful outcomes included correct classification of stalking by the police - instead of just harassment - as well as supporting and empowering victims (CVa16, Victim advocate); upskilling and raising awareness in other relevant agencies (CPo2, Police); and taking a holistic approach to risk management,

"We had a blank page to start with. Then that expertise that those individuals can bring in that multidisciplinary setting. We have got very senior practitioners across all of those disciplines who are able to bring a really expert perspective to address the problem" (SLT003, Central project management).

Based on our understanding of the IASU's workflow processes from the stakeholder interviews (see Fig. 2) and the literature on multi-agency partnerships dealing with high-risk cases (Robinson, 2004), we hypothesised that three possible causal mechanisms could bring about the first aim of MASIP, i.e. reduction in reoffending. The first of these was: improved classification of stalking cases and better investigation, aided by informed case review, would lead to appropriate charging decisions and higher convictions, thus ensuring the offender is restricted in reoffending for the time they are in custody or under court mandated restrictions. Steps on the path to increased convictions were identified by our interviewees as desirable interim outcomes, with knowledge and information sharing being one of the most important. As this interviewee said, "a lot of what we wanted to do as a unit was around spreading knowledge, sharing expertise, education". (CPo1, Police).

Stakeholders indicated their awareness of the fact that conviction is unlikely to address the obsessive fixation underlying stalking behaviour, however, stakeholders representing different agencies highlighted the importance of improved conviction for different reasons: police - because increased convictions would improve public confidence in the criminal justice system and because it restricts opportunity to commit offences while incarcerated; probation - because it provides them the opportunity to supervise and work closely with perpetrators; health - because it allows for the opportunity for health professionals to assess whether the perpetrator would benefit from health interventions and/or for the court to mandate conditions or treatment that might address the fixated behaviour, provided the perpetrator is willing to engage with treatment. Finally, victim advocates – because it reassures victims their complaint has been taken seriously and because it possibly restricts the offender from offending whilst incarcerated.

The second mechanism proposed was that bespoke health and other

[4] Comparator police agencies were collectively chosen by the project managing agency (Suzy Lamplugh Trust) and the evaluation team based on their similarity to the pilot agency. This similarity was judged via three criteria: 1) the comparator was in the 'most similar group' of police agencies, as statistically matched on demographic, social and economic characteristics, by His Majesty's Inspectorate of Constabulary and Fire and Rescue Services, see: https://hmicfrs.justiceinspectorates.gov.uk/police-forces/data/. 2) Stalking offences time-series trend trajectory, from data reported to the UK Home Office for collation. 3) If the comparator had a Police and Crime Commissioner (PCC) funded services (i.e., where stalking was mentioned as a priority in the PCC's strategy). In compliance with the data sharing agreements signed, we are obliged to anonymise the comparator police agency.

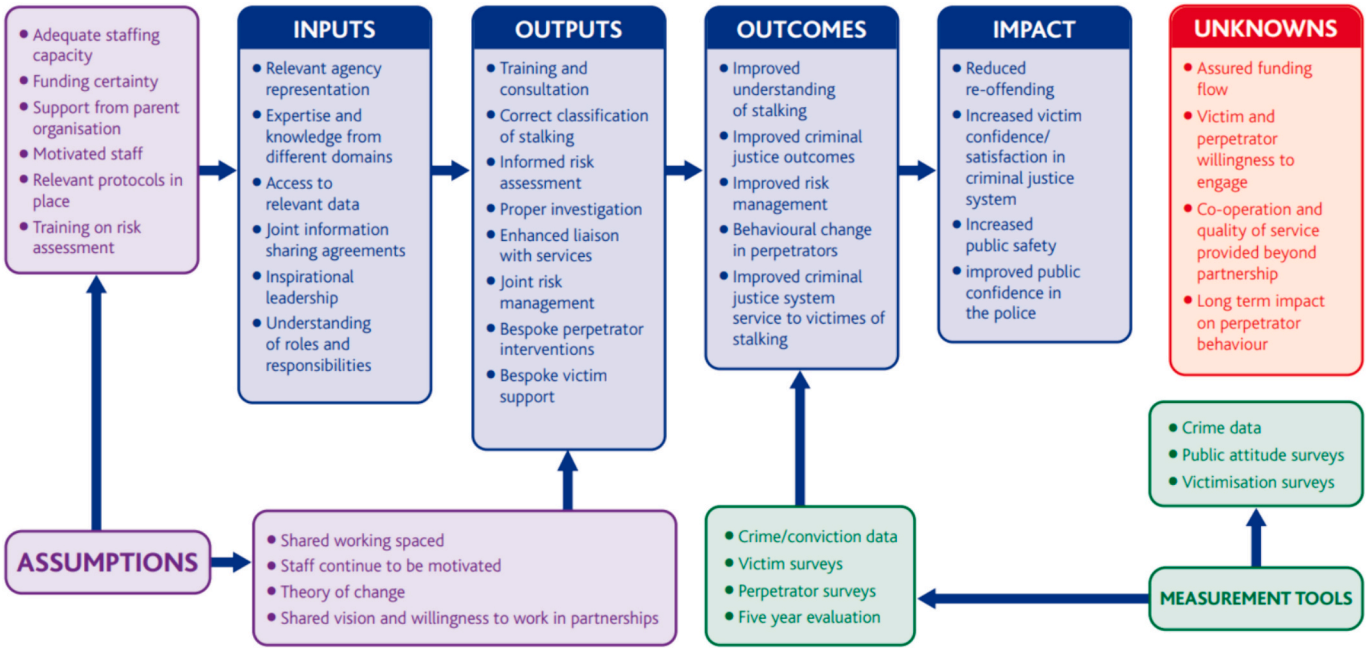

Fig. 1. Programme Theory of Change for IASU/MASIP.

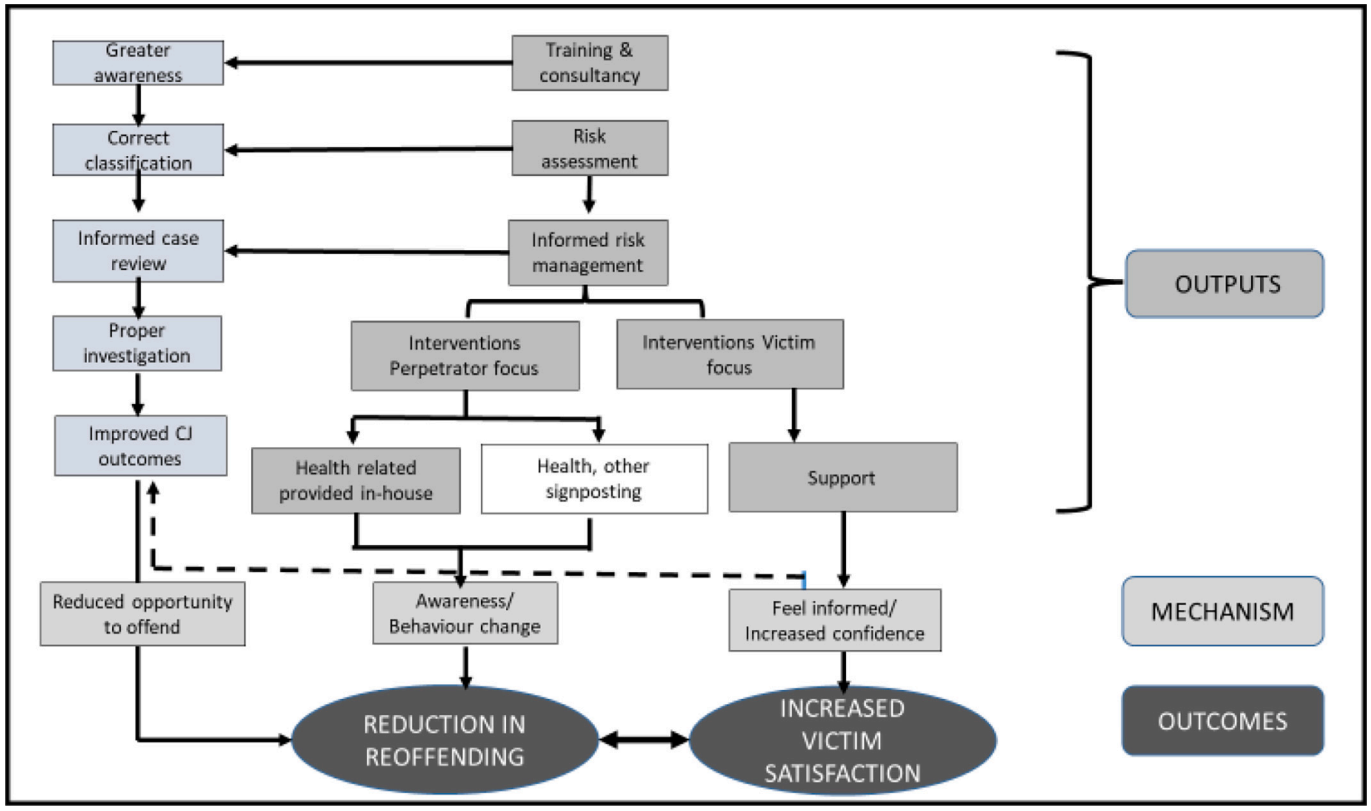

Fig. 2. Process Map for IASU/MASIP.

interventions designed to address the perpetrator's needs would lead to change in offender behaviour, ultimately leading to a reduction in reoffending. One of the biggest 'successes' of the project was the delivery of bespoke treatments designed for individual stalkers. As one stakeholder concluded,

"But even within types of stalkers, there is variance and differentiation and diversity… you need to understand what's driving the behaviour for that individual at that time, rather than thinking, 'Oh well, they're this type of stalker, so therefore this is the motivation.' That's not the case. And also, therapeutically, we've had to adapt and modify how we've worked with individuals, because of their competencies and lack of competencies, and skills and lack of skills, and underlying intelligence levels; we've had to adapt and accommodate and modify because of their learning styles, which is right." (CH3, Health).

Finally, the third mechanism suggested was that multi-agency working, by its very nature of collaborative information sharing will lead to a more effective and informed risk management plan and its execution, thus leading to reduction in opportunities to reoffend by being able to see both sides of the interaction. As one interviewee said,

"So, working in the unit where we have had information being shared from both the perpetrator and victim allows professionals to understand the whole story, or a lot of the story compared to what both individuals are saying. It just keeps adding pieces to that jigsaw, and you're starting to be able to see what's happening and being curious to find out what those missing pieces are" (CH15, Health/Outreach).

Knowledge sharing and multi-agency working also helped individual stakeholders change their perspective from being focused on their individual agency outcomes to a wider understanding, as this interviewee said, "I've gone from being very victim focused to being holistically focused". (CVa16, Victim Advocate). Furthermore, as one interviewee from the central project management team explained the advantages of multi-agency working was,

"I think there is also something about having an awareness of our differences and letting go of that preciousness about it, really. Being able to recognise that there is more than one way of resolving all of these issues and that nobody is necessarily right or wrong." (SLT001, Project Team).

It was thus clear that some of the articulated aims of the project were but essential steps in achieving the two ultimate goals of reducing reoffending and improving victim satisfaction.[5]

7.2. Was the IASU effective at reducing reoffending?

Over the data period September 2018 – December 2019 the partnership classified 103 stalking cases as those that they believed that they could add value to. This referred to whether the unit would provide some actual support to either the investigation or the risk management process as well as cases where the team had the capacity and capability to provide support, therapy, or services inhouse (when the perpetrator was not already in treatment and preliminary work had established that treatment was clinically indicated[6]). Given that each case produced a large amount of casework, this is, as is the norm in crime prevention evaluations, a modest sample (Eck, 2006).

The first mechanism proposed to reduce reoffending was higher convictions of perpetrators, which was tested by comparing the charge and caution rates for Cheshire with a comparator police force in the period August 2016–December 2019. Charges and cautions are not convictions but were believed to be an indicator of whether the mechanism was tracking in the right direction (e.g., more appropriate criminal justice outcomes). Detailed results are published elsewhere (Tompson, Belur, & Jerath, 2020) but, in brief, this revealed that Cheshire (IASU) brought charges in 17.7% of stalking cases, compared to 14.6% in the comparator force. Yet, Cheshire were more likely to charge for less serious forms of stalking, as defined by sentencing policies (known as 2 A offences, with the Home Office code of 195/12) more often, and more serious forms of stalking (4 A offences, with codes 008/65 and 008/66) less often than their comparator force. At face value, this does not seem like a success. However, Cheshire were far less likely to encounter evidential difficulties than the comparator force, suggesting that the investigations themselves were more successful. Overall, this provides a mixed picture of whether this first mechanism was operating as intended. But it should be noted that charges take a considerable amount of time to eventuate, and the data period only captured the beginning phase of MASIP. Hence, this premature evaluation of IASU (necessitated by the funders, the UK Home Office) may underestimate the true effects of IASU's long-term goal to increase convictions of perpetrators. The impact of continuing austerity in E&W and uncertain funding streams has led to funders being unable to commit to longer term evaluations. A development that might be self-defeating in the long run in terms of producing any substantial or conclusive evidence.

To test the second mechanism, whereby reoffending is hypothesised to be reduced by the bespoke health and other interventions delivered by IASU we turned to the partnership data.[7] Of the 71 cases where six months or more had elapsed since the unit began monitoring the case, there was evidence that 30 stalkers, not all of whom received a health intervention, continued their stalking behaviour (known as persistence in the clinical literature).[8] This equated to 42.3% of the sample reoffending. As mentioned previously, this is likely an inflation of the true reoffending rate because this is optimally calculated from when the health intervention concludes, and this date was not captured by IASU in the tight project timescales. The reoffending rate presented is lower than the 60% persistence rate found by Mohandie et al. (2006) which was the most comparable study based on sample type (community based), intervention type (threat assessment and legal) and included ex-intimate partners in the sample (see Appendix). Although with a small sample such as the one here it is naturally prudent to advise caution on the reliability of these results. The full breakdown of reoffending can be seen in Table 1.[9] Visual inspection of this shows that 20.4% of stalkers had reoffended by three months; 42.3% by six months; 47.7% by nine months and 58.3% by twelve months. Of the 44 cases where more than six and less than nine months had elapsed since the IASU had been monitoring the case, there was evidence that stalking behaviour had recurred (i.e., after a break of 6 months) to the *same* victim in 4.5% of cases ($n = 2$). This is lower than the range that is reported in research that has studied recurrence following a psychological intervention (9.5% - 40%, see Appendix), and in these two cases perpetrators had not received a health (psychological) intervention from the IASU.

To consider the impact of the third mechanism – that multi-agency working would lead to more effective risk management – we conducted separate analysis on the reoffending outcome data. Table 2 provides a breakdown of the 71 cases where enough time had elapsed to judge

Table 1
Quarterly breakdown of reoffending

	< 3 months		< 6 months		< 9 months		< 12 months	
Reoffended	n	%	n	%	n	%	n	%
Not reached time threshold	10		32		59		79	
No	74	79.6	41	57.7	23	52.3	10	41.7
Yes	19	20.4	30	42.3	21	47.7	14	58.3
Total	103		103		103		103	

[5] The latter goal - victim satisfaction - is not the focus of this study, but interested readers can find two further mechanisms related to this objective in Jerath et al., 2022.

[6] The sample of cases that IASU worked on was not known to differ from the wider population of cases in terms of stalker 'type' (e.g., using stalking typologies) or assessed risk level but may have differed in offender 'readiness' to engage in treatment.

[7] As police agencies outside of the pilot sites were not devoting additional resources to the accurate recording of stalking, we could not use these data to test research questions associated with reoffending.

[8] Despite our original intention to measure both stalking behaviour frequency and severity in the evaluation, the practitioners who collected the reoffending data preferred just to capture frequency. Hence, we cannot speak to the point that there can be decreases (and, indeed, increases) in harm even when frequency of stalking is unchanged.

[9] *Re*-offending was counted once only in Table 1. Put otherwise, if reoffending had (say) been recorded within the first and third quarters it would only count as one instance in Table 1 at the earliest occurrence.

Table 2
Intervention type by whether there was any evidence of reoffending

Intervention type	Yes	No	Total
Comprehensive risk assessment	4	6	10
Perpetrator direct intervention	3	14	17
Perpetrator health signposting	4	1	5
Perpetrator 3rd party consultation	21	26	47
Victim direct intervention	10	11	21
Victim target hardening	2	5	7
Victim 3rd party consultation	24	29	53

recidivism at six months post-IASU involvement, by the intervention type that was used in the case. These are not mutually exclusive categories. All victims received safeguarding, so this is not considered an intervention per se. Of those perpetrators who received a direct health intervention, 17.6% reoffended. This is at the low end of the range when compared to reoffending rates in studies where the perpetrators have received a health intervention with the measurement caveat expressed previously (0–52.9%, see Appendix), and suggests that the health interventions delivered in the pilot site were modestly successful.[10]

We present the results of the logistic regression in Table 3, with the coefficients converted into odds ratios to assist interpretation along with 95% confidence intervals. Visual inspection of Table 3 reveals that none of the independent variables, which represent various intensities of multi-agency working with perpetrators and victims, are statistically significant. Applying the Firth method confirmed these results.[11] Whilst disappointing, these results are unsurprising in the context of the sample being severely underpowered.

7.3. *What implementation features were crucial to the theory of change?*

From our qualitative fieldwork we identified factors necessary for the effective and efficient working of multi-agency partnerships and the successful achievement of intended outcomes. These are intended to guide models of partnership working like the one reported herein. Some of these, such as the importance of data sharing agreements, are specific to the E&W context and may not be applicable to partnerships elsewhere.

Facilitating factors are grouped into those considered essential preconditions and those regarded as essential for continued operational support for the IASU's ToC. Most factors are within the control of partnership agencies, but some are dependent on external factors, outside the remit of partnerships.

The first of three essential preconditions is full organisation and individual stakeholder commitment to the partnership work. Stakeholders interviewed insisted that all key players involved in the partnership ought to be motivated and willing to work as a team to be effective. As this interviewee said,

"It's the right people being involved. Not somebody being sent to a meeting because that's your job to go to meetings. We want people who are interested, passionate, willing to learn and engage and actually they see the benefit in what this does for them…They come along, galvanise the catalyst for change. These are the people that drive the cultural attitudes, the peers…" (CPo1, Police).

The willingness to work together is also predicated on support from senior leaders in the partner organisations. The setting up of the MASIP project augmented funding and resources to the already existing unit in Cheshire, thus gaining the backing of senior leaders in the respective organisations. However, stakeholders made it clear that substantial impetus and the leadership required for joint working came initially from individuals within relevant agencies. Furthermore, there was acceptance that without the support of the senior leadership team, partnership working would be very difficult.

Secondly, our observations indicated the importance of having information sharing agreements in place prior to or at the start of the partnership. These often take longer than expected and can act as an impediment to integrated working. Fortunately, IASU had these in place due to the pre-existing unit. This helped pave the way for smoother and more efficient partnership meetings and setting up databases to monitor the progress and outcome of cases over the course of the pilot.

Thirdly, the interview data indicated that joint training on stalking and how to work in partnerships would be useful. Partnership members require a shared language and shared understanding of the problem to conduct joint risk assessments. Stakeholders rated the two-day Stalking Risk Profile (SRP) training they received early in the project as crucial. As this interviewee said, "I think MASIP has also given us the opportunity to go on training we would never have gone on before." (CSs3, Social Service). When asked, stakeholders said they had never received training on partnership working, despite some of them working in partnerships for over 20 years. They reflected that they would have valued such training, including understanding each other's databases, key performance indicators, and chain of command.

At its core, adequate resourcing and time are essential for the project set up and operational phases. Stakeholders mentioned that recruiting and training suitable personnel took a considerable amount of time. Furthermore, individual stakeholders were often overworked, or needed more support in terms of resources. For example, petty cash needed to be made available so that the IASU victim advocate did not use personal funds for refreshments when meeting with victims in safe public spaces to gain their confidence. Sustained drain on specific agency or individual resources, financial or otherwise, can prove detrimental to the success of the partnership work. This is another area where inadequate funding might have a substantial impact on whether all the necessary conditions required for the intervention to work could be satisfied.

Secondly, co-location and a shared working space were highly valued by the IASU stakeholders so that partners could easily access relevant expertise from partner agencies and take joint decisions.

"There are definitely processes that developed along the way that we've all had to learn and actually being located in the same office space I think has been invaluable in making sure that everyone's roles are contributing to the same overall vision." (CH14, Health/Outreach).

Furthermore, the importance of building relations of trust between partners cannot be overstated. This invariably takes time. IASU had mature working processes whereby close working interpersonal relationships had been nurtured over time and appropriate governance structures and mechanisms were in place to resolve disagreements. As one stakeholder, external to the core IASU, but part of the wider multi-agency consultation said about the IASU,

"I think the partnership works quite well because you have no hierarchy in the meeting which I think is really good, there's no fear to speak or contradict, which I think is a really positive… I think there is

Table 3
Logistic regression; reoffending at 6 months as the dependent variable (n = 71), AIC: 102.53

	OR	s.e.	CI		*p*-value
Intercept	0.86	0.77	0.18	3.97	0.85
Perpetrator direct intervention	0.74	0.85	0.12	4.51	0.74
Perpetrator indirect intervention	0.96	0.81	0.20	5.01	0.96
Perpetrator combination intervention	1.94	0.91	0.37	10.98	0.44
Victim direct intervention	0.48	0.75	0.09	2.22	0.36
Victim indirect intervention	0.41	0.80	0.11	1.38	0.16
Victim combination intervention	2.20	0.63	0.52	10.46	0.30

[10] It is important to note that there were myriad reasons why direct health interventions were not always delivered. A perpetrator could be unreceptive to treatment, could be assessed and found not to be suitable for treatment, or could move out of the pilot area or become deceased.

[11] Results available on request from the corresponding author.

the freedom to disagree, there's the freedom to talk something through and I think by doing that you can come to a much better conclusion of what the risks are, how we can try and manage the risk." (CSs3, Social Service).

Another interviewee commented on the result of trusting relations built over time on the working of the IASU,

"We are integrated, we sit side by side, we share competency. When we sit around the table, we don't talk as psychologist, outreach practitioner, support office, and ISAC, we speak as a collective." (CPo4, Police).

And finally, the willingness of perpetrators and victims to work with the partnership is vital. In the absence of engagement from perpetrators and victims the partnership has little chance of effecting behavioural change. Trust and confidence from victims, especially, is not a given (Robinson, 2004). Similarly, stakeholders acknowledged that reducing the obsessive compulsive thought patterns that impelled stalking behaviour through therapeutic interventions was dependent on perpetrators wanting to engage with the process, as the team did not have any authority to compel them to engage with therapy.

8. Discussion

In this study we report a theory-driven evaluation of an English multi-agency partnership to prevent the harm associated with stalking. A ToC approach has several purposes from contributing to resourcing and planning, describing the process, aiding with monitoring and evaluation, and identifying feed-back loops (Stein & Valters, 2012). In our evaluation we found the ToC approach useful for three reasons: first, it helped guide practitioners to establish how programme activities, supported by appropriate resources, could be aligned with intended outcomes via clearly articulated mechanisms. Second, it supported a richer contribution to the evidence base as it enabled us to generate and test a priori hypotheses to better understand how or why interventions may work (or not). Last, and perhaps most importantly, it permitted an assessment of interim outcomes when long-term outcomes could not be observed within the circumscribed timelines of the multi-agency initiative. It is important to note that lack of funding for assessing the medium- and long-term impact of interventions is a de facto feature of evaluations in the public sector, which depend on supplemental government funding that is decided upon each financial year. We can only add our voice to the general plea to evaluation commissioners to ensure that the goal of evidence-based practice is supported by adequate funding for appropriate follow-up periods for evaluations.

The ToC elicited from stakeholders in the first phase of the process evaluation enabled the generation of research questions and hypotheses that were subsequently tested with qualitative and quantitative data. The evaluation's mixed methods approach adopted the EMMIE framework (Johnson et al., 2015) which prompts that we not only ask 'what works?', but also 'how it works, for whom, and in what circumstances?'. This requires attention to the mechanisms and the circumstances in which those mechanisms are activated, neutralised, or deactivated. We set out to answer three research questions: how was the IASU envisaged to work? Did it reduce reoffending)? And what implementation features were crucial to the operation of the ToC? The sample was too modest to be able to effectively examine for whom (stalker type, based on the relationship between the stalker and victim (McEwan et al., 2020)) it worked.

Clear visual depiction of a ToC can be a powerful "tool for describing the key components of an intervention and how it impacts on outcomes" (De Silva et al., 2014: 10), and it can help other researchers and policy makers understand how a complex intervention works. However, visually depicting complex multi-agency initiatives where several parts occur simultaneously or sequentially can be 'a mess of criss-crossing boxes and arrows' (Funnell & Rogers, 2011). The evaluation team worked with the stakeholders in presenting a simplified and linear process diagramme (Wilkinson, Hills, Penn, & Barbrook-Johnson, 2021) which linked different activities to outcomes (Fig. 2).

The ToC enabled the identification of several mechanisms to achieve the intended outcomes of reducing reoffending. We subsequently used the insight provided by the ToC to test hypotheses with quantitative data. This would have enabled us to evaluate both outcomes and causal mechanisms (Breuer, Lee, De Silva, & Lund, 2015), had the requisite data been collected. We specified the data trends we expected to see if the mechanisms were working as intended in advance of the quantitative data analysis. (Belur, Tompson, & Jerath, 2019). This guarded against loss of independence (i.e., working closely with the stakeholders) and subjectively selecting results that biased the findings of the evaluation. The fact that we were unable to test all the hypotheses due to either insufficient data being collected or the short duration of the evaluation in no way reduces the value of a ToC approach to evaluating such a multi-strategy initiative.

Overall, the hypotheses regarding whether the IASU partnership reduced reoffending are not strongly supported. We found only partial support for the hypothesis that improved investigation would lead to higher charges (as an interim outcome for convictions) in stalking cases. However, given that demonstrating evidence of effect would have been extremely challenging in the short evaluation period, the reoffending rates emanating from the IASU indicate that there is cause for optimism. For example, 17.6% of all perpetrators receiving a psychological intervention reoffended by six months, which compares favourably to commensurate rates reported in studies of other psychological interventions, albeit with usually forensic samples (see Appendix). The broader reoffending rate (for all perpetrators) was 42.3% at six months, which again falls within the range reported in other studies (see Appendix). Given that measurement of reoffending began at the point the case came to the attention of the IASU – rather than at the completion of treatment as used in other studies - it is likely that the IASU reoffending rates are inflated in relation to comparators. This is further compounded by the measure of stalking reoffending being more comprehensive than other studies, due to the excellent communication between the victim advocate and the victims. Furthermore, we accept that reoffending exists along a continuum of severity, but the reoffending data collected was not reflective of that continuum, in that, it did not distinguish between high risk and high harm reoffending from less serious reoffending, making it difficult to make a nuanced assessment of whether the interventions had an impact on the level of risk to victims. Overall, it would not be inaccurate to claim that the effects of IASU in reducing reoffending are likely underestimated here given that the reporting rate was perhaps better than in comparable studies.

It would be worth flagging up that there were several aspects to the intervention, including raising awareness, helping with better investigation, providing bespoke victim support for risk management and throughout the criminal justice process, and better offender management, apart from providing bespoke therapy where clinically indicated. The IASU looked at all stalking related cases and provided some inputs throughout the investigation and supported the prosecution of these cases if they were charged. Thus, some parts of the intervention were applicable to all perpetrators and all stalking cases. As far as bespoke therapy was concerned - some perpetrators were selected for therapy where it was either clinically indicated, or where the team made a judgement about whether the perpetrator would benefit from the bespoke therapy provided and focused on those cases where, in their opinion they could "add value". In the absence of further details about how they made that decision, our conclusion would be that the programme is suitable for all cases in achieving the various outcomes (better registration, investigation, successful prosecutions, reduction in victim fear of crime, as well as reduced reoffending). However, bespoke therapy might be useful in cases where it is clinically indicated and where the perpetrator is willing to engage.

Crime reduction interventions often fail to have the intended effect, but it is not always clear whether this is attributable to theory or implementation failure. Implementation failure is a potent threat in

social programmes that depend on a protracted and complicated chain of events to occur before an intervention can be considered completed (Bowers et al., 2018). Our evaluation helped identify implementation issues relevant to the ToC which included serious commitment and willingness from senior leadership in organisations and stakeholders to work together; information sharing agreements in place; and joint training to facilitate partnership work. Conditions to support smooth partnership working included adequate resourcing, co-location where possible, adequate time to build trust between partners and, finally, willingness of perpetrators and victims to engage with the partnership. None of these factors are new with respect to multi-agency working (Cleaver et al., 2019; Davies, 2021; Davies, Barlow, & Fish, 2023; Solomon, 2019).

8.1. The pros and cons of a ToC driven evaluation

The literature reports several key challenges with taking a ToC evaluation approach. Here, we discuss some of these challenges and how they were addressed in our evaluation. The foremost challenge arises from the fact that complex initiatives often have vague designs. Consensus of long-term outcomes can be straightforward; however, stakeholders can struggle to specify interim outcomes and activities and how they fit in a causal process (Connell & Kubisch, 1998). Despite the IASU having a good operational grasp of how their partnership intended to achieve its aims, we experienced this challenge in articulating causal processes.

Compounding this, it is not unusual for partner stakeholders to have conflicting ideas of what outputs might best lead to desirable outcomes. For example, if the overall goal of an initiative is to reduce reoffending, the police may approach this through more detection and incarceration, while probation and health may think that receiving appropriate treatment would lead to desistance. Agreeing on a ToC, with shared outcomes and impact can be difficult in partnerships with multiple stakeholders. However, given the fact that the IASU already had built trust between partners, partnership working was made easier because of shared goals within the team.

The process of designing the ToC to the satisfaction of all partners was iterative and intensive, and involved a workshop with stakeholders and myriad other fieldwork activities (e.g., observations, interviews). Hence, it was not merely a 'box ticking exercise' as De Silva et al. (2014) caution against and as Dyson and Todd (2010) note, it was labour and resource intensive. This was true for both the evaluation team and the stakeholders, which some might feel diminishes resources for the implementation of the initiative (Cole, 2003). Yet, the benefits of adopting a ToC approach manifest when stakeholders are intensively consulted in its design and have a deeper level of engagement with the intervention itself (De Silva et al., 2014). We contend it also helped the construction of a shared professional identity, resolving what has previously been identified as a crisis of professional identity experienced by stakeholders in multi-agency working (Souhami, 2019). We suggest that the co-construction of a ToC was critical to evidencing the budding success of the IASU.

One drawback of the close collaboration needed for this evaluation might be that the evaluation team influenced the delivery of the ToC itself. Hence, it could be argued that we became participants in the evaluation through our dialogue with stakeholders and the interim findings we reported. Whilst this negates controlling for externalities in the evaluation, as might be preferred by an experimental approach, it allowed us to document contextual conditions and to collect data that best suited the purposes of the evaluation, rather than retrofitting this at the end of the programme. Furthermore, the 'collaborative consensus building approach' helped create a 'synergistic ToC' where the focus was not so much on judging only whether the intervention worked, but on the 'shared learning journey' as part of a less traditional evaluation approach (Laing, 2022).

Conceptualising ToCs visually as two-dimensional figures means that they are necessarily simplified. On the one hand, this can mean they are easier to understand but, on the other, the complexities and interactions between the different moving parts of a complex social intervention are not captured and 'may conceal some of the explanatory power of the causal pathways' (Breuer et al., 2015). The latter can lead to some interdependencies and interactions being overlooked. Alternatively, a complex ToC with many nested components and feedback loops can make the measurement of interim outcomes within the model and subsequent evaluation problematic (De Silva et al., 2014). Accordingly, we chose to include top level strategic interim outcomes for the quantitative analysis of the hypotheses and used qualitative evidence to support other outcomes.

In sum, adopting a ToC driven approach to evaluating the MASIP and IASU proved to be advantageous on grounds identified by Breuer et al. (2015). First, because this approach is compatible with any evaluation framework, including the realist inspired EMMIE framework used herein. Second, for complex social interventions it is an alternative to randomised control trials, where such trials are difficult to implement. Finally, describing the process by which the ToC was framed, how the evaluation was designed, and pathways to change measured, allows for the judgement of validity of the findings (Dyson & Todd, 2010) and provides a roadmap for practitioners and policy makers to set up future partnerships.

9. Conclusion

This evaluation of IASU to reduce the harmful effects of stalking indicated good reason to believe that the multi-agency effort was promising with regards to preventing reoffending. We say only promising, because the evaluation ran parallel to the initiative and had no follow up period. Hence, we were unable to ascertain whether the ambitious long-term outcomes were achieved. Instead, the ToC approach helped to identify where causal processes were functional and leading in the desired direction. We believe a longer follow up period and a larger sample size are essential to demonstrate with some confidence the impact of complex interventions delivered by partnerships like the IASU. Although our attempt to test these theoretically generated hypotheses was limited by the availability of data, we believe that a ToC driven approach not only helps identify potential mechanisms and provides guidance on identifying interim and final outcomes and how to measure them, but it also helps practitioners identify how or why there are different strands of the intervention and how various outputs (activities undertaken by stakeholders) should complement each other. Furthermore, a ToC driven approach combined with an EMMIE inspired evaluation is valuable as it can contribute towards specifically identifying whether interventions work, how they work, for whom they work, and under what conditions. At the very least, our attempt (through its limitations) demonstrates the importance of data collection and provision for long term evaluation plans, especially for complex social interventions.

CRediT authorship contribution statement

Jyoti Belur: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Lisa Tompson:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Kritika Jerath:** Writing – review & editing, Writing – original draft, Validation, Investigation, Formal analysis, Data curation.

Declaration of competing interest

None.

Acknowledgement

Funding was procured by the Suzy Lamplugh Trust from the Home Office's Police Transformation Fund (2018/19) and subcontracted to UCL for the evaluation that this study stems from.

Appendix A. Appendix

Table X
Summary of re-offending statistics from academic research, with type of re-offending measure. N.B. recidivism means re-offending over a non-specified period, persistence refers to an ongoing offence to the same victim within 6 months and persistence refers to a break in stalking behaviour at least six months before resuming (to the same or a different victim). These definitions are taken from McEwan et al. (2009)

Study	Study country	Sample size	How measured	*Re*-offending rate	Re-offending type
Eke et al., 2011	North America	78	(Sub-sample of Mohandie et al., 2006). Police contacts after threat assessment and legal intervention.	56% ($n = 44$)	Recidivism
Foellmi et al., 2016	North America	89	Evidence of post-assessment stalking behaviour or harassment after psychological intervention from self-reports and criminal records.	35% ($n = 31$)	Recidivism
Hehemann et al., 2017	The Netherlands	115	Report of stalking made to police within six months of index offence.	40% ($n = 46$)	Persistence
James et al., 2010	England	222	When a subject re-contacted their target (no ex-intimates in sample). No timeframe in English sample who received a legal warning. Australian sample received a psychological intervention.	44.1% ($n = 98$)	Persistence
	Australia	140		50% ($n = 70$)	Persistence over 12 weeks
				27.1% ($n = 38$)	Persistence over 52 weeks
Malsch et al., 2011	The Netherlands	709	New convictions for people already convicted of stalking, over 4 years or more.	11% (n missing)	Recidivism (stalking)
				24% (n missing)	Recidivism (stalking related)
McEwan et al., 2009	Australia	200	Prior stalking, to the same or other victim, disclosed during assessment with mental health clinician during a psychological intervention.	33% ($n = 65$)	Recidivism
				9.5% ($n = 19$)	Recurrence
McEwan et al., 2017	Australia	157	Stalking prior to index offence was taken from criminal records and clinician reports.	10% ($n = 16$)	Persistence/ Recurrence (same victim)
				27% ($n = 43$)	Recurrence (different victim)
McEwan et al., 2018	Australia	235	Stalking charge or restraining order (or similar), to the same or other victim, after psychological treatment. Over an average follow up time of 4 years. 8 offenders stalked both.	14.9% ($n = 35$)	Recurrence - same victim
				13.6% ($n = 32$)	Recurrence - new victim
				26.4% ($n = 62$)	Recidivism
				40% ($n = 28$)	Recurrence – same victim
McEwan et al., 2020	The Netherlands	70	Police reports by the same or other victim, over an average follow up time of 2.63 years from referral to psychological intervention. 6 offenders stalked both.	21.4% ($n = 15$)	Recurrence – new victim
				52.9% ($n = 37$)	Recidivism
Mohandie et al., 2006	North America	730	When a subject re-contacted their target, subsequent to threat assessment and legal intervention.	60% ($n = 434$)	Persistence
Rosenfeld, 2003	North America	189	Evidence of a second arrest or re-contact of target (based on victim or Probation reports or indicated in clinical interview).	49% ($n = 93$)	Recidivism in whole sample
Rosenfeld et al., 2007	North America	29	Police and/or probation reports for 20 months after a psychological intervention.	0% ($n = 0/14$)	Recidivism for treated stalkers
				26.7% (n = 4/15)	Recidivism for treatment dropouts
Shea, 2015	Australia	148	Police charges for stalking or indicative stalking post index offence and post psychological intervention. Sample is cases not offenders.	20.3% ($n = 30$)	Recidivism (stalking charge)
				30.4% ($n = 45$)	Recidivism (indicative stalking charge)
Shea et al., 2018	Australia	146	Stalking post index offence and post psychological intervention was taken from criminal records and self-reports.	33.6% ($n = 49$)	Recurrence

References

Atkinson, M., Jones, M., & Lamont, E. (2007). *Multi-agency working and its implications for practice* (pp. 28–42). Reading: CfBT Education Trust.

Barnes, M., Matka, E., & Sullivan, H. (2003). Evidence, understanding and complexity: Evaluation in non-linear systems. *Evaluation, 9*(3), 265–284.

Belur, J., Tompson, L., & Jerath, K. (2019). *MASIP evaluation interim report*. UCL Discovery. https://discovery.ucl.ac.uk/id/eprint/10078857/1/MASIP%20evaluation%20interim%20report%20v2.6.pdf.

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. H. Office (Ed.). *Research Report, 52*.

Björklund, K., Häkkänen-Nyholm, H., Sheridan, L., & Roberts, K. (2010). The prevalence of stalking among finnish university students. *Journal of Interpersonal Violence, 25*(4), 684–698. https://doi.org/10.1177/0886260509334405

Blamey, A., & Mackenzie, M. (2007). Theories of change and realistic evaluation: Peas in a pod or apples and oranges? *Evaluation, 13*(4), 439–455.

Bowers, K., Gill, P., Morgan, R., Meiklejohn, S., & Johnson, S. D. (2018). Challenges for EMMIE as a realist evaluation framework. In *Realist Evaluation for Crime Science* (pp. 98–118). Routledge.

Breuer, E., Lee, L., De Silva, M., & Lund, C. (2015). Using theory of change to design and evaluate public health interventions: A systematic review. *Implementation Science, 11*, 1–17.

Burnett, R., & Appleton, C. (2004). Joined-up services to tackle youth crime: A case-study in England. *British Journal of Criminology, 44*(1), 34–54.

Chapman, J., Higgins, A., & Hales, G. (2017). Slough site report. In *Police Effectiveness in a Changing World*. Police Foundation. Available at: http://www.police-foundation.org.uk/publication/police-effectiveness-changing-world-slough-site-report/.

Cleaver, K., Maras, P., Oram, C., & McCallum, K. (2019). A review of UK based multi-agency approaches to early intervention in domestic abuse: Lessons to be learnt from existing evaluation studies. *Aggression and Violent Behavior, 46*, 140–155.

Cole, M. (2003). The health action zone initiative: Lessons from Plymouth. *Local Government Studies, 29*(3), 99–117.

Connell, J. P., & Kubisch, A. C. (1998). Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects, and problems. *New approaches to evaluating community initiatives, 2*(15–44), 1–16.

Cox, C. (2019). Policing a diverse society: The community based rationale for multi-agency working. In *Multi-agency working in criminal justice* (pp. 105–120). Policy Press.

Davies, P. (2021). Practicing co-produced research: Tackling domestic abuse through innovative multi-agency partnership working. *Crime Prevention and Community Safety, 23*(3), 233–251.

Davies, P., Barlow, C., & Fish, R. (2023). The hard and complex work of implementing new multi-agency risk assessment approaches to policing domestic abuse. *Crime Prevention and Community Safety, 25*, 148–165.

De Silva, M. J., Breuer, E., Lee, L., Asher, L., Chowdhary, N., Lund, C., & Patel, V. (2014). Theory of change: A theory-driven approach to enhance the Medical Research Council's framework for complex interventions. *Trials, 15*(1), 1–13.

Dyson, A., & Todd, L. (2010). Dealing with complexity: Theory of change evaluation and the full service extended schools initiative. *International Journal of Research and Method in Education, 33*(2), 119–134.

Eck, J. E. (2006). When is a bologna sandwich better than sex? A defense of small-n case study evaluations. *Journal of Experimental Criminology, 2*, 345–362.

Eke, A., Hilton, N., Meloy, J., Mohandie, K., & Williams, J. (2011). Predictors of Recidivism by Stalkers: A Nine-year Follow-up of Police Contacts. *Behavioral Sciences & the Law, 29*, 271–283. https://doi.org/10.1002/bsl.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika, 80*(1), 27–38.

Foellmi, M. C., Rosenfeld, B., & Galietta, M. (2016). Assessing risk for recidivism in individuals convicted of stalking offenses: Predictive validity of the guidelines for stalking assessment and management. *Criminal Justice and Behavior, 43*(5), 600–616.

Funnell, S., & Rogers, P. (2011). *Purposeful programme theory: Effective use of theories of change and logic models*. San Francisco: Jossey-Bass.

Gray, P. (2014). 'Report of Research on Youth Offending Team Partnerships and the Social Context of Youth Crime', Plymouth Law and Criminal Justice Review, 6. Available at: https://pearl.plymouth.ac.uk/handle/10026.1/9007. http://hdl.handle.net/10026.1/9007.

Hehemann, K., van Nobelen, D., Brandt, C., & McEwan, T. (2017). The reliability and predictive validity of the Screening Assessment for Stalking and Harassment (SASH). *Journal of Threat Assessment and Management, 4*, 164–177.

Her Majesty''s Inspectorate of Constabulary Fire and Rescue Services and Her Majesty''s Crown Prosecution Service Inspectorate. (2017). *Living in fear—The police and CPS response to harassment and stalking. A joint inspection by HMIC and HMCPSI*. London, England: Her Majesty's Inspectorate of Constabulary Fire and Rescue Services. Retrieved from http://www.justiceinspectorates.gov.uk/hmic/publications/living-in-fear-the-police-and-cps-response-to-harassment-and-stalking/.

Herbert, J. L., & Bromfield, L. (2019). Better together? A review of evidence for multi-disciplinary teams responding to physical and sexual child abuse. *Trauma, Violence, and Abuse, 20*(2), 214–228.

Hester, M., Eisenstadt, N., Jones, C., & Morgan, K. (2017). *Evaluation of the drive project - a pilot to address high-risk perpetrators of domestic abuse year 1 feasibility study*. University of Bristol.

Holt, T. J., Lee, J. R., Liggett, R., Holt, K. M., & Bossler, A. (2019). Examining perceptions of online harassment among constables in E&W. *International Journal of Cybersecurity Intelligence and Cybercrime, 2*(1), 24–39.

Horwath, J., & Morrison, T. (2007). Collaboration, integration and change in Children's services: Critical issues and key ingredients. *Child Abuse and Neglect, 31*, 55–69.

Jamal, F., Fletcher, A., Shackleton, N., Elbourne, D., Viner, R., & Bonell, C. (2015). The three stages of building and testing mid-level theories in a realist RCT: A theoretical and methodological case-example. *Trials, 16*(1), 1–10.

James, D. V., Kerrigan, T. R., Forfar, R., Farnham, F. R., & Preston, L. F. (2010). The fixated threat assessment centre: Preventing harm and facilitating care. *Journal of Forensic Psychiatry and Psychology, 21*(4), 521–536. https://doi.org/10.1080/14789941003596981

Jerath, K., Tompson, L., & Belur, J. (2022). Risk management in stalking victims: a multi-agency approach to victim advocacy. *Journal of interpersonal violence, 37*(11–12), NP8989–NP9015.

Jerath, K., Tompson, L., & Belur, J. (2023). Treating and managing stalking offenders: findings from a multi-agency clinical intervention. *Psychology, Crime & Law, 29*(10), 1161–1184.

Johnson, S. D., Tilley, N., & Bowers, K. J. (2015). Introducing EMMIE: An evidence rating scale to encourage mixed-method crime prevention synthesis reviews. *Journal of Experimental Criminology, 11*(3), 459–473.

Kelly, R. (2020). The problematic development of the stalking protection order. *Modern Law Review., 83*(2), 406–427.

Korkodeilou, J. (2016). Stalking victims, victims of sexual violence and criminal justice system responses: Is there a difference or just 'business as usual'? *British Journal of Criminology, 56*(2), 256–273. https://doi.org/10.1093/bjc/azv054

Lea, S. J., & Callaghan, L. (2016). "It Gave Me My Life Back" An Evaluation of a Specialist Legal Domestic Abuse Service. *Violence Against Women, 22*(6), 704–721.

Laing, K. (2022). The contribution of a 'synergic theory of change' approach to democratising evaluation. *Research for All, 6*((1), 8), 1–17.

MacKenzie, R. D., & James, D. V. (2011). Management and treatment of stalkers: Problems, options, and solutions. *Behavioral Sciences & the Law, 29*(2), 220–239.

MacKenzie, R. D., McEwan, T. E., Pathé, M. T., James, D. V., Ogloff, J. R. P., & Mullen, P. E. (2009). *Stalking risk profile: Guidelines for the assessment and management of stalkers* (1st ed.). Melbourne, Victoria, Australia: StalkInc. and Centre for Forensic Behavioural Science, Monash University.

Malsch, M., de Keijser, & Debets, S. E. C. (2011). Are stalkers recidivists? Repeated offending by convicted stalkers. *Violence and Victims, 26*(1), 3–15. https://doi.org/10.1891/0886-6708.26.1.3

Mazerolle, L., Cherney, A., Eggins, E., Hine, L., & Higginson, A. (2021). Multiagency programs with police as a partner for reducing radicalisation to violence. *Campbell Systematic Reviews, 17*(2), Article e1162.

McEwan, T. E., Harder, L., Brandt, C., & de Vogel, V. (2020). Risk factors for stalking recidivism in a Dutch community forensic mental health sample. *International Journal of Forensic Mental Health, 19*(2), 127–141.

McEwan, T. E., Mullen, P. E., & MacKenzie, R. (2009). A study of the predictors of persistence in stalking situations. *Law and Human Behavior, 33*, 149–158.

McEwan, T. E., Shea, D. E., Daffern, M., MacKenzie, R. D., Ogloff, J. R. P., & Mullen, P. E. (2018). The Reliability and Predictive Validity of the Stalking Risk Profile. *Assessment, 25*(2), 259–276. https://doi.org/10.1177/1073191116653470

McEwan, T. E., Shea, D. E., Nazarewicz, J., & Senkans, S. (2017). Reassessing the link between stalking and intimate partner abuse. *Partner Abuse, 8*(3), 223–250.

McKeon, B., McEwan, T. E., & Luebbers, S. (2015). "It's not really stalking if you know the person": Measuring community attitudes that normalize, justify and minimise stalking. *Psychiatry, Psychology and Law, 22*(2), 291–306.

McNamara, C. L., & Marsil, D. F. (2012). The prevalence of stalking among college students: The disparity between researcher- and self-identified victimization. *Journal of American College Health, 60*(2), 168–174. https://doi.org/10.1080/07448481.2011.584335

Mohandie, K., Meloy, J., McGowan, M. G., & Williams, J. (2006). The RECON typology of stalking: Reliability and validity based upon a large sample of North American stalkers. *Journal of Forensic Sciences, 51*(1), 147–155. https://doi.org/10.1111/j.1556-4029.2005.00030.x

Morgan, S., Pullon, S., Garrett, S., & McKinlay, E. (2019). Interagency collaborative care for young people with complex needs: Front-line staff perspectives. *Health and social care in the community, 27*(4), 1019–1030.

Mullen, P. E., Pathé, M., & Purcell, R. (2000). *Stalkers and their victims*. Cambridge University Press.

Nijdam-Jones, A., Rosenfeld, B., Gerbrandij, J., Quick, E., & Galietta, M. (2018). Psychopathology of stalking offenders: Examining the clinical, demographic, and stalking characteristics of a community-based sample. *Criminal Justice and Behaviour, 45*(5), 712–731. https://doi.org/10.1177/0093854818760643

Ostermeyer, B., Friedman, S. H., Sorrentino, R., & Booth, B. D. (2016). Stalking and violence. *Psychiatric Clinics, 39*(4), 663–673.

Owens, J. G. (2016). Why definitions matter: Stalking victimization in the United States. *Journal of Interpersonal Violence, 31*(12), 2196–2226.

Pathé, M. T., Haworth, D. J., Goodwin, T. A., Holman, A. G., Amos, S. J., Winterbourne, P., & Day, L. (2018). Establishing a joint agency response to the threat of lone-actor grievance-fuelled violence. *The Journal of Forensic Psychiatry and Psychology, 29*(1), 37–52.

Pawson, R., & Tilley, N. (1997). An introduction to scientific realist evaluation. In *Evaluation for the 21st century: A handbook, 1997* (pp. 405–418).

Reeves, C. (2013). How multi-agency are multi-agency risk assessment committees? *Probation Journal, 60*(1), 40–55.

Robinson, A. L. (2004). *Domestic violence MARACs (multi-agency risk assessment conferences) for very high-risk victims in Cardiff*. Wales: A Process and Outcome Evaluation. Cardiff University School of Social Sciences.

Robinson, A. L., & Clancy, A. (2021). Systematically identifying and prioritising domestic abuse perpetrators for targeted intervention. *Criminology and Criminal Justice, 21*(5), 687–704.

Rogers, P. (2014). Theory of change. *Methodological briefs: Impact evaluation, 2*(16), 1–14.

Rosenbaum, D. P. (2002). Evaluating multi-agency anti-crime partnerships: Theory, design, and measurement issues. *Crime Prevention Studies, 14*, 171–225.

Rosenfeld, B., Galietta, M., Ivanoff, A., Garcia-Mansilla, A., Martinez, R., Fava, J., … Green, D. (2007). Dialectical behavior therapy for the treatment of stalking offenders. *International Journal of Forensic Mental Health, 6*(2), 95–103. https://doi.org/10.1080/14999013.2007.10471254

Rosenfeld, B. (2003). Recidivism in stalking and obsessional harassment. *Law and Human Behavior, 27*(3), 251–265. https://doi.org/10.1023/A:1023479706822

Shea, D. E. (2015). *The reliability and predictive validity of the Guidelines for Stalking Assessment and Management (SAM). PhD thesis*. Monash University.

Shea, D. E., McEwan, T. E., Strand, S. J. M., & Ogloff, J. R. P. (2018). The reliability and predictive validity of the guidelines for stalking assessment and management (SAM). *Psychological Assessment, 30*(11), 1409–1420. https://doi.org/10.1037/pas0000589

Sloper, P. (2004). Facilitators and barriers for co-ordinated multi-agency services. *Child: Care, Health and Development, 30*(6), 571–580.

Solomon, M. (2019). Becoming comfortable with chaos: Making collaborative multi-agency working work. *Emotional and Behavioural Difficulties, 24*(4), 391–404.

Souhami, A. (2019). Multi-agency practice and professional identity. In *Critical practice with children and young people 2nd edition* (p. 179).

Stein, D., & Valters, C. (2012). Understanding theory of change in international development. In *Justice and security research programme paper 1*. London: London School of Economics and Political Science. https://eprints.lse.ac.uk/56359/1/JSRP_Paper1_Understanding_theory_of_change_in_international_development_Stein_Valters_2012.pdf.

Storey, J. E., & Hart, S. D. (2011). How do police respond to stalking? An examination of the risk management strategies and tactics used in a specialized anti-stalking law enforcement unit. *Journal of Police and Criminal Psychology, 26*, 128–142.

Tompson, L., Belur, J., & Jerath, K. (2020). *MASIP evaluation final report.* UCL Discovery https://discovery.ucl.ac.uk/id/eprint/10097009/.

Tompson, L., Belur, J., & Jerath, K. (2021). A victim-centred cost–benefit analysis of a stalking prevention programme. *Crime Science, 10*, 1–11.

Van Tulder, R., & Keen, N. (2018). Capturing collaborative challenges: Designing complexity-sensitive theories of change for cross-sector partnerships. *Journal of Business Ethics, 150*(2), 315–332.

Vander Laenen, F., & Persac, N. (2014). Whose claim is legitimate anyway? Negotiating power in inter-agency collaboration. *Legitimacy and Trust in Criminal Law, Policy and Justice: Norms, Procedures, Outcomes*, 111–129.

Walby, S. (2005). Improving the statistics on violence against women. *Statistical Journal of the United Nations Economic Commission for Europe, 22*(3-4), 193–216.

Weiss, C. H. (1997). How can theory-based evaluation make greater headway? *Evaluation Review, 21*(4), 501–524.

Wheatley, R. C. (2019). *What drives men who commit stalking offences and how practitioners can best respond to their needs*. United Kingdom: Nottingham Trent University.

Wilkinson, H., Hills, D., Penn, A., & Barbrook-Johnson, P. (2021). Building a system-based theory of change using participatory systems mapping. *Evaluation, 7*(1), 80–101.