BMJ Open

Searching for the mechanisms of change: a protocol for a realist review of batterer treatment programmes

Alisa J Velonis,[1,2] Rebecca Cheff,[1] Debbie Finn,[1] Whitney Davloor,[1] Patricia O'Campo[1,3]

CrossMark

[1]Centre for Research on Inner City Health, St Michael's Hospital, Toronto, Ontario, Canada
[2]Division of Community Health Sciences, University of Illinois at Chicago, School of Public Health, Chicago, Illinois, USA
[3]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

Correspondence to
Dr Alisa J Velonis; avelonis@uic.edu

ABSTRACT

Introduction: Conflicting results reported by evaluations of typical batterer intervention programmes leave many judicial officials and policymakers uncertain about the best way to respond to domestic violence, and whether to recommend and fund these programmes. Traditional evaluations and systematic reviews tend to focus predominantly on whether the programmes 'worked' (eg, reduced recidivism) often at the exclusion of understanding for whom they may or may not have worked, under what circumstances, and why.

Methods and analysis: We are undertaking a realist review of the batterer treatment programme literature with the aim of addressing this gap. Keeping with the goals of realist review, our primary aims are to identify the theory that underlies these programmes, highlight the mechanisms that trigger changes in participant behaviour and finally explain why these programmes help some individuals reduce their use of violence and under what conditions they are effective or not effective. We begin by describing the process of perpetrator treatment, and by proposing an initial theoretical model of behaviour change that will be tested by our review. We then describe the criteria for inclusion of an evaluation into the review, the search strategy we will use to identify the studies, and the plan for data extraction and analysis.

Ethics and dissemination: The results of this review will be written up using the RAMESES Guidelines for Realist Synthesis, and disseminated through peer-reviewed publications aimed at the practitioner community as well as presented at community forums, and at violence against women conferences. Ethics approval was not needed.

Strengths and limitations of this study

- Realist syntheses are based on the development of a solid initial theory describing how, why, for whom and under what conditions programme strategies generate key outcomes.
- We present our initial theory for a realist synthesis of batterer intervention programme evaluations, and our explanation for the importance of conducting such a review.
- Our initial theory draws from existing theory about batterer intervention programmes, and presents a hypothesis about how the primary strategies of education, skills building and group process might generate immediate outcomes including participants' desire to develop alternatives to violence, use of non-violent communication and violence avoidance skills, development of empathy for partners and shifts among others.
- We present a search strategy and approach to analysis that are consistent with realist principles.

debates have emerged, highlighting points of contention about the nature of intimate partner violence (IPV), about the multiple levels and types of influences that may contribute to abusive behaviour, about what 'success' means in terms of batterer treatment and about the 'right' approach to both achieving and evaluating this success.[1 4] These differences in perspective lead to conflicting conclusions about the effectiveness of these programmes, making the job of navigating the literature that surrounds batterer intervention or treatment programmes challenging.

Yet, as part of the official response to domestic violence across North America, guidelines on sentencing, or other codified judicial requirements frequently require individuals who are convicted of crimes against intimate partners to attend treatment or educational programmes (the content and format of which can vary widely) as a condition to receiving a deferred sentence, probation or parole.[5–7] This, among other factors, has

INTRODUCTION

Since the emergence of early batterer intervention or treatment programmes (BIPs) in the 1970s and 1980s, discussions about their efficacy have proliferated within research, professional and policy circles.[1–3] In attempting to answer the general question, 'Do these programmes work?' a number of subsurface

spurred a proliferation of programme evaluations and systematic reviews, most sharing a defined goal of determining whether or not the programmes that currently exist can be proven to directly reduce subsequent violence and criminal behaviours.[3 8] While there has been significant debate over what theoretical approach(es) should be used to guide these programmes (eg, feminist theory, family systems theory, cognitive behavioural theory), the nature of these discussions tends to be as political (eg, profeminist or antifeminist in rhetoric) as it is scientific.[9–12] Few systematic reviews have attempted to examine the underlying programmatic theory and understand how, why and in what contexts these programmes work, or do not work.

Focusing on theory rather than on programmes

Drawing on the work of Sayer[13] and other realist philosophers, Pawson[14] describes interventions as 'complex process that are inserted into complex structures' (p.79). Budgets get cut, referrals increase or decrease, participants resent attending mandated programmes, staff feel overworked and underappreciated, and the list goes on. Even under the best of conditions, behaviour change is difficult, and programmes often use multiple theoretical strategies to help clients move along the path to improvement.

Intervention programmes for batterers are particularly tricky. While multiple models for batterer intervention exist, most function within the framework of a larger community and criminal justice-oriented response to domestic violence.[1] Participants are primarily—although not exclusively—required to attend BIPs as part of probationary or deferred sentencing agreements,[6] and while the programmatic details vary across jurisdictions, most BIPs are designed as a series of educational and skill-building group sessions that run from 12 to 52 weeks.[7] Regardless of the specific therapeutic, philosophical or political framework used, these programmes are impacted by a variety of internal and external factors, including the characteristics and experiences of participants and staff, the mission of the lead organisation, the levels of communication between the programmes, the local courts or probationary departments, victim-centred domestic violence services, and the social and political climate of the larger community. In turn, these factors influence how programmes run, how closely they adhere to the programme design, how the strategies are received by participants and more.

Yet, a great deal of the research describing the 'effectiveness' of batterer intervention programmes has been designed to minimise the influence of these real-world contextual factors, generally by controlling for many of the very forces that could explain programme success or failure (eg, cultural backgrounds, income, substance use). Often, experimental and quasiexperimental evaluations are held up as the gold standard; these models endeavour to link the intervention—and ONLY the intervention—to a narrowly defined outcome, usually

recidivism or reoffense.[3 8 15 16] A result has been the proliferation of a vast body of literature that shows mixed evidence of programme success with little explanation of why. It was our frustration with this lack of explanation (and the frustration that programme administrators and domestic violence advocates expressed to us about not knowing what to do to improve treatment) that led us to conduct this realist synthesis of the literature.

What follows is a description of a protocol we have developed for undertaking this review. As with other realist reviews, the purpose of this synthesis is explanatory: to articulate underlying programme theories and use evidence to determine their usefulness and relevance for batterer treatment.[14] In writing and sharing this protocol, we set forth three goals. First, based on our initial understanding of BIPs, to propose and explain a 'preliminary rough theory' that captures the framework used by the majority of BIPs with the aim of identifying both strengths and gaps. Second, we wish to illustrate how a realist perspective can add value to our understanding and interpretation of BIP evaluation by shifting focus from whether programmes work to how and why they (should) work. Finally, we desire to provide transparency for the forthcoming review, enabling readers to know the specific steps that will be followed throughout the review process and to highlight some of the specific challenges we face as we move forward.

Part of what makes realist synthesis unique is the emphasis it places on proposing, testing and ultimately refining theory. Initially, reviewers conceptualise a rough, preliminary theory that explains 'what is supposed to happen?', and 'why is that supposed to work?'[17] for the programme in question (a programme theory). In the language of realist synthesis, the key elements that need to be identified and understood in relation to one another include the programme strategies (the activities that comprise the programme), contextual factors that influence how and for whom the programmes operate (such as the individual-level characteristics mentioned above as well as the structural context in which programmes operate, such as funding and statutory requirements), participant outcomes (eg, participants reduce their use of violence or recidivism) and the hidden mechanisms or 'generative processes' that often take place within participants' minds and trigger the outcome (or, as Wong et al (ref. 17, p.6) put it, 'what it is about a programme that generates change').[17–19]

For each intervention strategy that a programme includes, there may be multiple pathways that lead to multiple outcomes; each pathway has its own set of contextual factors and mechanisms at play, and the outcomes can be visible or hidden, intermediate or final, and intended or unintended.[20] At the end of a realist synthesis, the primary goal is the generation of a refined theory that takes the shape of a set of context–mechanism–outcome (CMOs) configurations, a heuristic used to illustrate these relationships and pathways.[20] It is at this point that our theoretical lens becomes less focused

on the programmatic elements of BIPs and more focused on the mechanisms of change.

This protocol paper is organised into several sections following the steps of realist synthesis laid out by Pawson et al[19] (figure 1). The core of any realist synthesis or evaluation is the description of the programmatic theory that underpins the activity being evaluated (Step 1). We began this process by formulating our preliminary (rough) theory describing the strategies most BIPs employ, the circumstances in which they operate, what they aim to accomplish and how it appears to us that these strategies will lead to those outcomes; essentially, this is a programme theory that describes, in general terms, what is supposed to change and why as a result of the programme. As we develop this protocol, we are in the initial stages of the review process, and what is reflected here reflects our initial thinking about BIPs. Furthermore, as Jagosh and colleagues remind us, realist reviews are abductive in nature, meaning that we infer 'to the best explanation', iteratively 'examining evidence and developing hunches or ideas about the causal factors linked to that evidence' (ref. 20, p135). After laying out our preliminary programme theory, we go on to describe the remaining steps (2–5) that will be taken to conduct this review. Finally, we conclude with a discussion of why we believe this approach will contribute to our understanding of batterer treatment programmes, and how it can inform not only programme implementation but also larger policy.

METHODS/DESIGN

The impetus for this review was a series of conversations with programme facilitators and judicial personnel who had grown frustrated with the lack of conclusive information about what their systems could do to reduce the perpetuation of partner violence. As they saw it, the evidence was insufficient to declare batterer interventions useless, yet these programmes clearly did not work in all circumstances. Using this to inform our approach, we framed our research question as for whom, and under what conditions, will batterer intervention programmes help men who have been identified as perpetrators of partner violence reduce their violent behaviour and why?

Step 1a: establishing the scope of our work

Before refining our scope and articulating the processes we believe to be at work, we needed to understand what
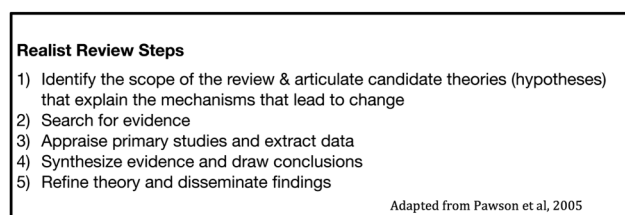
Realist Review Steps

1) Identify the scope of the review & articulate candidate theories (hypotheses) that explain the mechanisms that lead to change
2) Search for evidence
3) Appraise primary studies and extract data
4) Synthesize evidence and draw conclusions
5) Refine theory and disseminate findings

Adapted from Pawson et al, 2005

Figure 1 Key steps in realist review.

the bigger picture of batterer response looked like in North America. After selecting a handful of evaluations, systematic reviews and reports from both the scientific and grey literature,[2 3 5 8 15 21–27] we learned that the majority of individuals who attend batterer intervention in the USA and Canada undergo a two-part process: first, they enter the criminal justice system and are adjudicated for an offense against a partner; and second, they are mandated to attend an educational/therapeutic 'treatment' programme—what we refer to as a BIP—as part of their sentence or agreement with the court. Since the vast majority of participants in BIPs first have contact with the criminal justice system (and because our interest is in explaining how and why these programmes work), we consider this to be part of the larger environment in which these programmes exist, rather that as one of the strategies or interventions that make up BIP programmes. We recognise that this decision may become a limitation for this study, and that to gain a true understanding of the mechanisms that underlie BIPs, we may need to look more critically at the system-level issues at work (eg, whether or not the level of communication between BIP programme staff and court officials has an impact on BIP outcomes); if this appears to be the case as we proceed, we will revisit this decision. However, the ongoing debate over the efficacy of the programmes themselves (and the lack of information about community-level systems contained within most programme evaluations) leads us to begin this review by focusing on BIPs.

Another issue we encountered while considering the scope of the review was the enormity of what we were calling 'individual perpetrator characteristics', and how they may interact with programme strategies to impact whether participants respond to the programme (positively or negatively). A substantial amount of literature supports the contention that men who engage in violence against intimate partners and family members are not all alike.[28] Over the past 30 years, researchers have attempted to categorise batterers according to psychological, behavioural, attitudinal and/or motivational characteristics using descriptors such as family only or typical batterers, dysphoric/borderline or passive-aggressive-dependent batterers, and sociopathic or generally violent/antisocial batterers.[29–32] Likewise, while Johnson does not provide a batterer typology, per se, he differentiates between types of IPV (coercive-controlling violence, situational partner violence and violent resistance), which suggests that the motivation behind these categories of violence would be different for different instigators.[33] Finally, substantial evidence points to an overlap between substance abuse, a history of trauma, neglect and/or victimisation, and other psychological conditions, some of which are captured in the batterer typologies described above, but which may also emerge as independent issues that programmes may or may not be prepared to address.[34 35]

In no way are we claiming that individual-level characteristics cause someone to engage in abusive

behaviours; rather, those factors may interfere with the effectiveness of programme strategies.[36][37] Unfortunately, individual programmes are often unable to assess for and address the myriad of issues that participants may bring to the intervention, and even in jurisdictions where completing an intake assessment for substance abuse and mental health problems is mandatory, the availability of and coordination between various treatment modalities can vary widely. Understandably, much of this detail is left out of evaluation write-ups and formal reports, yet these are influential factors that need to be acknowledged and addressed. For the purposes of this review, we have decided that we will consider them to be among the contextual factors that can influence programme effectiveness, and where they are mentioned, we will note them accordingly. However, because of the wide breadth of possible influences, and the relative dearth of information collected about them in most evaluations, we have chosen not to limit our review to only those evaluations that address these issues.

As we began developing our rough programme theory for BIPs, we first drafted a flow chart illustrating the 'big picture' processes at work in most perpetrator interventions (figure 2). After reviewing various programme descriptions and evaluations, we differentiated two sets of outcomes that were generally discussed: 'proximal' outcomes, or changes that happen within the participants as a result of participation in the programme activities (such as changes in attitudes, skills and intentions); and 'final' outcomes, which would include recidivism and reassault, and which are generally measured as long-term consequences of programme participation.

The majority of batterer intervention evaluations are concerned primarily or exclusively with recidivism and/or reassault, which we found problematic for several reasons. First, we contend that for BIPs to impact recidivism or reassault, they first must achieve these more immediate changes in attitudes, motivations and skills. The final outcomes are at least one step removed from the BIP itself in that any changes in recidivism or violent behaviour that are not preceded by changes in attitudes, motivations and skills may not be the result of the BIP programme. Furthermore, even after the proximal outcomes are met (if and when they are met), numerous factors unrelated to the BIP programme itself may influence whether participants reoffend. For example, if a perpetrator with co-occurring substance abuse or clinical depression receives little or no additional treatment for those problems (whether because he refuses treatment or because treatment is not available, affordable or accessible), the progress he makes towards realising a non-abusive relationship (that may be achieved through BIP participation) may be offset by the lack of assistance for these other problems.

Another problem with using recidivism or reoffense measurements as the sole indicators of success or failure of BIPs is that definitions of these outcomes often differ across studies and may not be limited to assaults against a partner, but can include a conviction for any violent crime, an arrest for a violent crime (regardless of conviction) or even any subsequent arrest.[1] Finally, because many—if not most—acts of abuse do not result in law enforcement intervention, using this as an outcome likely underestimates the recurrence of these behaviours.[1][3][38] Unfortunately, other measures of reassault are likely to be equally unreliable, in that unless evaluators contact the perpetrators' current partners, these data rely primarily on participants' self-reported descriptions of their behaviours. Even when victim reports are included, the numbers are often small and subject to self-report bias. For these reasons, we chose to limit our review to the relationship between BIPs and these proximal outcomes, which are addressed in greater detail below.

Step 1b: forming an initial theory of change: looking for the mechanisms

Owing to the lack of agreement in the scientific community about whether partner violence is primarily a psychological, cognitive, developmental or social problem, the literature is replete with a variety of intervention approaches. One of the challenges we faced was in identifying a universal programme model that reflects all or most BIPs; not only do different state and provincial jurisdictions outline different standards for programme length and content,[6][7] but even at a local level, programmes can vary tremendously in how they are implemented and who they serve. In spite of this heterogeneity, the vast majority of programmes across the USA and Canada employ what could be called a feminist-informed cognitive-behavioural therapy (CBT)[i] model.[10][28][39] While the specific content, philosophical emphases and details differ across jurisdictions and programmes, our initial review of programme descriptions indicate that most BIPs include a common set of elements, including: (1) the use of educational strategies that challenge beliefs about gender equity, relationships and the impact of abuse; (2) skill-building activities intended to provide alternatives to abuse and violence; and (3) a facilitated group process that offers participants both support and accountability.[40]

Based on this understanding of 'generic' BIPs, the review team constructed our initial (rough) programmatic theory. After identifying what we felt were the most essential programme-level (proximal) outcomes—those that would be necessary to lead to further change in longer term (final) outcomes like recidivism—we worked backwards, asking ourselves what strategies were likely to be linked to each outcome, and how that

---

[i]Although this approach is often referred to as 'the Duluth Model' because of the influential community-based programme that emerged from that city in the early 1990s, we are referring to it as a feminist-informed CBT model because it is likely that very few current programmes are true replications of the Pence and Paymar's Duluth programme.[16]
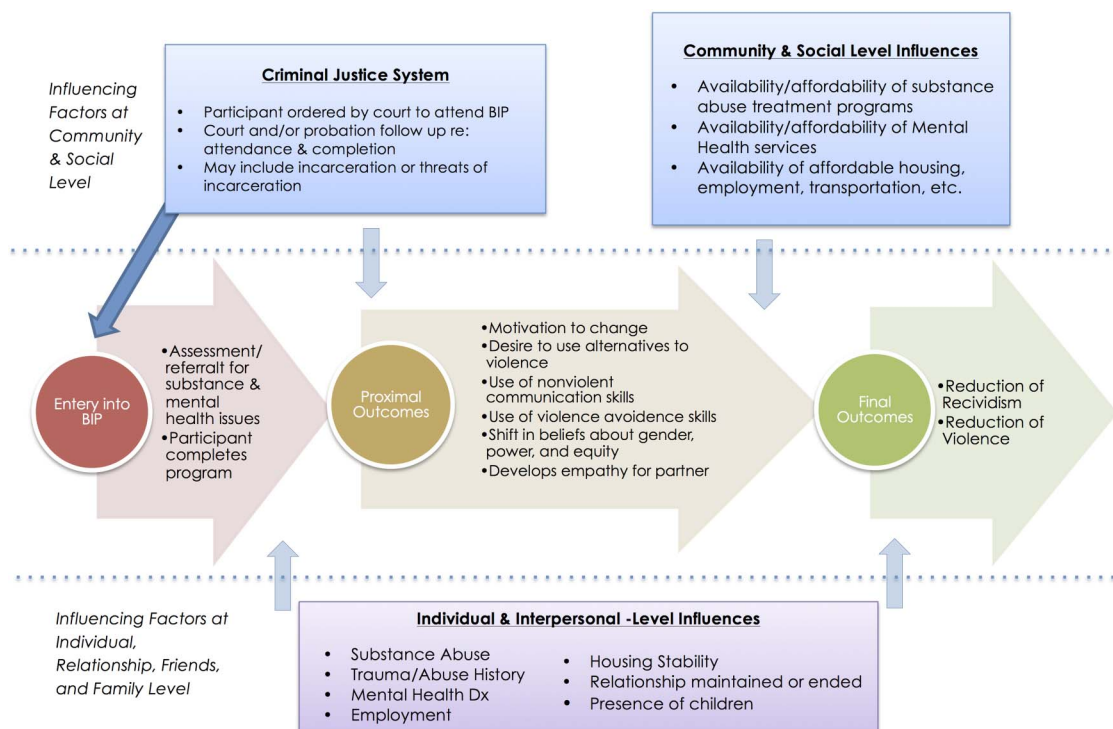
Figure 2 Batterer intervention process, outcomes and influencing factors.

strategy triggers the outcome and what contexts are the most relevant in allowing that to happen? We recognised that each strategy may be linked to multiple outcomes, that multiple mechanisms are often at work within each strategy, and that sometimes change has to happen in a particular order. For example, one 'education strategy'
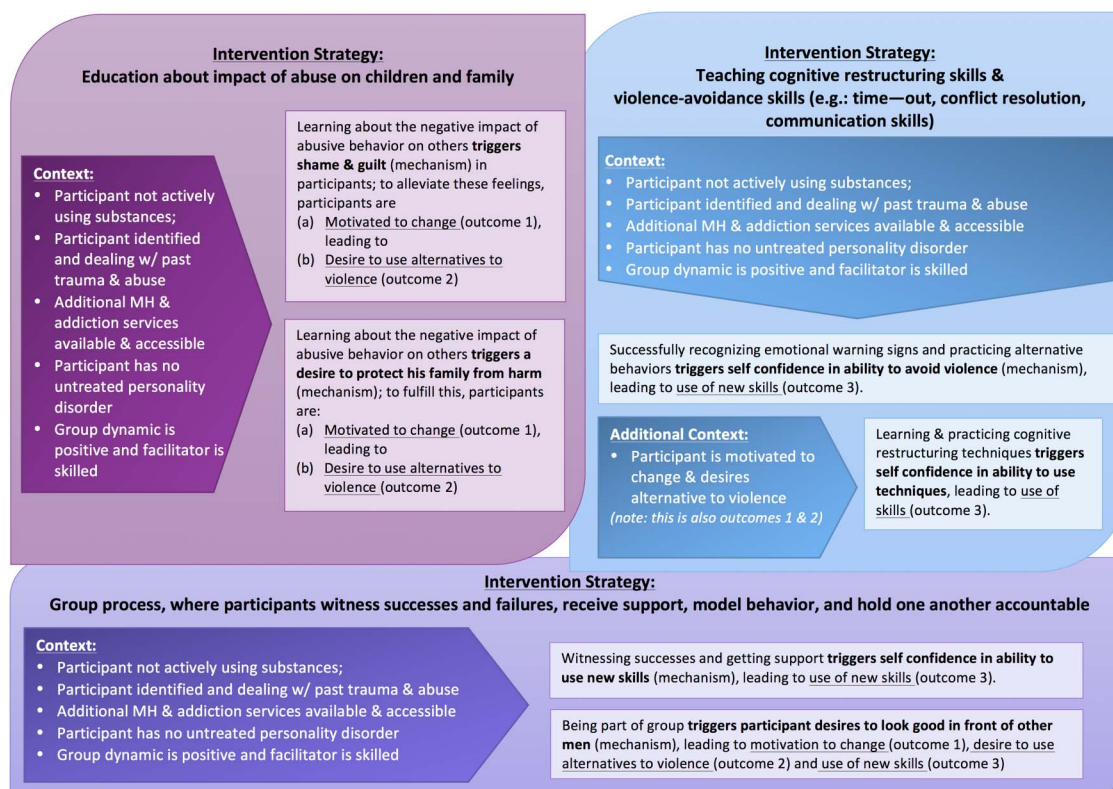

Figure 3 Sample context–mechanism–outcome configurations from the preliminary rough theory of batterer intervention treatment programmes.

that appears common to most BIPs is to discuss the negative impact that violence and abuse has on one's partner and children. For participants who are capable of feeling remorse and empathy (ie, who do not have sociopathic tendencies or an antisocial personality disorder) (context), learning about these damaging impacts triggers both shame and guilt for past behaviours (mechanism), and a desire to stop hurting people whom he loves (mechanism). Both these mechanisms can lead to the perpetrator feeling motivated to stop using abusive behaviours (outcome) and to desire alternatives to violence (outcome). Another strategy commonly employed is to teach perpetrators skills they can use to avoid becoming violent (such as recognising emotional triggers and calmly walking away). Learning, and then practising, these behaviours triggers a level of self-confidence in participants (mechanism) that leads to the eventual adoption of these skills (outcome). However, before a participant is likely to truly benefit from these skill-building sessions, he most likely needs to already feel motivated to stop using abusive behaviour and to desire alternatives to violence. Figure 3 illustrates several pathways in which programme strategies may link to our proximal outcomes; this is a partial model intended to exemplify our process, and is by no means a complete outline of our preliminary, rough theory.

At this stage, we wish to acknowledge that we fully anticipate revising and refining this as we proceed with the review. In their reflection on realist review, Jagosh and colleagues describe their struggle with identifying a singular theoretical construct that guided the subject of their review, and observed the ways in which context, mechanisms and outcomes often overlap, with an outcome in one chain of evidence serving as a context in a subsequent one.[20] We also recognise that using proximal outcomes may pose a challenge, as the nature of these outcomes—especially motivations and attitudes—can be subjective in nature and difficult to capture, and some may not be captured at all. As we move forward with our review, we will assess how well these constructs are assessed, and where we believe critical gaps may exist.

As we constructed our protocol, we identified several existing theories of behaviour change (eg, social cognitive theory, stages of change, etc) that appear useful for understanding how BIPs are supposed to work,[25 41] yet our attachment to these is preliminary. Through ongoing discussions within the team and continuously asking what causes a particular strategy to lead to a particular outcome during the synthesis phase of the review, we will continue to refine our models and flesh out the mechanisms at work.

Step 2: search for evidence
Search strategy
In partnership with a medical reference librarian (who is conducting the searches but is not part of the review team), we selected search terms that have been identified

in previous literature searches (examples include variations on words such as 'batterer', 'perpetrator', 'intervention' 'evaluation', etc). The disciplinary and interdisciplinary databases in our scope include (but are not limited to): MEDLINE, EBM Reviews (including Cochrane Database of Systematic Reviews), EMBASE, PsycINFO, CINAHL, Criminal Justice Abstracts, Social Sciences Abstracts, International Bibliography of the Social Sciences (IBSS), Applied Social Sciences Index and Abstracts (ASSIA), ProQuest Criminal Justice, ProQuest Dissertations & Theses Full Text, ProQuest Social Services Abstracts and Sociological Abstracts. Additionally, databases and electronic resources such as the Minnesota Center Against Violence and Abuse (MINCAVA) Electronic Clearinghouse and Google will be used to search the 'grey literature' for unpublished and informal evaluations. Based on prior experience with realist reviews as well as the literature describing them, we expect that as we proceed with the review, we will return to this step several times as we expand and/or refine our scope as necessary. All searches will be limited to English language articles published from 1995 to present.

Inclusion/exclusion criteria
Unlike more traditional syntheses, we are less concerned with whether or not an evaluation meets certain methodological standards (eg, is a randomised trial or includes a control/treatment group design) and more with the type of information it can provide about how, why and for whom BIPs work; this is reflected in our initial inclusion and exclusion criteria. Quantitative, qualitative and mixed-methods evaluations, regardless of study design, are to be included if the programmes they assess:

A. *Are offered in community-based rather than institutionalised settings, such as the military or prison.* Even if the programme activities and format resemble those found in community-based settings, BIPs conducted in these institutions likely involve a unique set of contextual factors and mechanisms that impact how and for whom the programmes work (eg, a soldier's commanding officer often ensures compliance with programme requirements, unlike in civilian settings)[3];

B. *Include some form of facilitated group treatment/education component*, the most common format required across North America. Most importantly, for purposes of this review, we are intentionally excluding research involving couples counselling or individual psychotherapy, as these are sufficiently different from BIP programmes, are considered controversial by many practitioners, and are often specifically prohibited by judicial statute[7];

C. *Run at least 8 weeks or 16 h in duration.* Most judicial statutes require at least 12 h,[7 24] and we do not believe that programmes that are shorter could be comparable in scope or content;

D. *Involve primarily male, court-mandated partner violence perpetrators.* Both research and observation suggest

that the use of violence may be different for women than for men, necessitating different approaches.[42] Likewise, men who voluntarily choose to participate in perpetrator treatment likely have different characteristics, and programmes designed to cater primarily to this population may be addressing different causes of violence;

E. *Measure at least one proximal outcome*, such as skills, attitudes, intentions. This criterion emerged as we developed our preliminary programme theory (described above), and reviewed articles identified during an early search. At the start of the project, we planned to include evaluations that measured at least one of the final outcomes, but the iterative process of theory identification led us to shift to a more limited scope.

Articles that are programme descriptions or evaluative reviews will be set aside for use as background, but will not be included in the formal synthesis process. Articles that focus solely on reducing programme attrition, increasing completion rates, or are limited to identifying the impact that participant's stage of change has on programme success will also be excluded.

Article screening

Using the search strategies outlined above, our librarian will generate a list of articles and (when available) abstracts, which we will divide among members of the review team, who will review the titles and abstracts to determine if the paper is (a) focused on domestic violence perpetrator programmes at all, and if it appears to (b) fit within the aforementioned inclusion/exclusion criteria. Screeners will be asked to categorise each article as 'include', 'exclude', and 'maybe'. When an abstract is not available, titles will be used to determine whether the article is appropriate for the review (eg, does the title mention BIPs?); if a title is insufficient to make this determination, the article will remain in the list of potential evaluations to be included until the complete text can be reviewed.

To ensure inner rater reliability, we will randomly select a handful of titles and abstracts that all screeners will review; as a group, we will discuss each screener's categorisation and, as necessary, come to consensus about articles in which screeners disagree. Once we are satisfied that all screeners share an understanding of the criteria and screening objectives, each team member will complete her/his assignments.

After this initial screening phase, all articles labelled as 'include' and 'maybe' will be redistributed among the review team members, who will complete a second screen of the remaining titles and abstracts. Once all members completed this task, the review team will again discuss this process. Articles that the first screener labels as 'include' but the second screener decides to 'exclude' will be discussed, and consensus reached.

Finally, the complete article or paper will be obtained for all remaining titles. Once again, inner rater reliability

will be assessed by having all reviewers read the same set of five articles, individually make recommendations on exclusion or inclusion, and then meet as a group to discuss the process. Each remaining article will be distributed among the review team members, and skimmed in order to make a final determination of whether to include or exclude it based on our screening criteria. If screeners are uncertain about whether or not to include a particular article at this stage, the article will be shared among other team members, and consensus will be reached.

Step 3: study appraisal and data extraction

For the appraisal process, each article will be read carefully by two reviewers, each assessing the relevance of the document to our inquiry (ie, how much information can it contribute to our development of programme theory?) and the rigour (ie, whether that information was generated using credible and trustworthy methods).[43] Reviewers will use a tool designed to identify and record the following information: (1) the programme strategies that are described, (2) what proximal outcomes are measured, (3) how the proximal outcome(s) is/are measured, (4) the contextual factors that are mentioned in the article (eg, if participants with addictions need to be in treatment or recovery prior to joining the programme), (5) whether the authors describe possible mechanisms that could lead to the outcome(s), and if so, what those mechanism are; and (6) the study design/fit for purpose making a clear note if it seems to bias the results (eg, if only successful programme participants were included in the evaluation). Based on these findings, the reviews will give an overall impression of the richness of the data available from this article and how much it can contribute to our understanding of programme theory.

We expect additional articles will be excluded after this in-depth review process if it is decided that they cannot contribute to our understanding of BIPs. If the two individuals who review a single article come to different conclusions, the larger team will discuss the issues and, if necessary, others will be asked to review the article(s) as well. Finally, we will comb through the citations of our articles as well as through our initial search results for additional articles or reports describing the same programme, and will review these sets or 'families' of articles as a single unit.

As reviewers read and re-read these papers, particularly pertinent passages will be directly extracted and included in the spreadsheet, and other data summarisd and annotated as necessary. The reviewers will meet on a regular basis to discuss their findings.

Step 4: Analysis and synthesis

Having started with a hypothesised theoretical model for BIPs, this stage will involve examining the evidence gathered and determining whether it supports or contradicts our proposed programmatic theory. Using our proximal

outcomes as our organising framework, we will look carefully at each evaluation that pertains to a particular outcome (or, if multiple publications describe the same programme, we will look at these as a 'family' of articles) and will assess how the data that were extracted from the studies inform our understanding of how batterer intervention works. Specifically, we will use the data to construct CMOs for each programme. Owing to the emphasis that many evaluators have placed on looking at final, rather than proximal, outcomes, and at whether BIPs lead to reductions in recidivism/reoffense (instead of how they lead to them), we anticipate that the data-describing mechanisms that underlay BIPs may be thin; thus, we will apply abductive reasoning as necessary to formulate our series of CMOs.[20] We will be particularly cognisant of the ways in which different contextual factors—when addressed—appear to influence the mechanisms that lead to these proximal outcomes.

Relying on the interdisciplinary perspectives and expertise of our research team, we will look at the information that arises from the construction of each CMO, as well as across programmes, to identify similarities and differences. Ultimately, we anticipate using the synthesis process to refine our original theoretical model in light of our review findings.

Step 5: Presentation and dissemination
The findings from this process will be presented in at least two formats: through at least one peer-reviewed article that conforms to the RAMESES publication standards put forth by Wong et al[43] that is intended to inform implementation scientists and others in academic settings; and through targeted outreach and conversations intended to reach decision-makers and practitioners: programme directors, policymakers, and community coalitions charged with overseeing coordinated responses to partner violence. This will include presentations at domestic violence and batterer treatment coalitions and conferences, and the preparation of plain-language reports and briefs for dissemination through national and international networks.

DISCUSSION
This review was conceived after informal conversations with batterer intervention treatment providers, domestic violence advocates and judicial system personnel revealed frustration over the current understanding of batterer intervention programmes. Results from evaluations and systematic reviews have been found to be conflicting, inconclusive and without substantial guidance about how approaches to batterer treatment could be improved.[2 44] In response, our team decided to complete this realist synthesis of perpetrator treatment programme evaluations with the aim of clarifying how and why BIPs work for some men, and the role that certain contextual factors may play in that success (or lack of success).

This is the first realist review of the perpetrator treatment evaluation literature that we know of, and we believe it will offer key insights into the debate over how communities can respond to partner violence. By focusing on the mechanisms that lead participants to change (rather than only looking at whether or not change was achieved), we believe we will provide much needed insight into promising (and not-so-promising) theoretically informed strategies that can lead to the types of changes that will allow men to reduce their use of violent and/or controlling behaviours.

One of the key contributions of this review, in relation to the majority of BIP programme evaluations and systematic reviews that have been done, is that we will focus on the impact that programmes have on proximal outcomes, rather than on reoffending or recidivism. We believe this is important for several reasons. To truly understand why programmes are or are not successful (and for whom and under what conditions), we need to gain a clear picture of the processes that lead to these outcomes. In the case of batterer treatment, it is unlikely that participants are magically transformed into non-violent partners simply because they attended a programme; rather, the programme promotes certain outcomes within participants that then lead to these more distal behavioural changes. Identifying what these proximal outcomes are and how programmes can achieve them is a key part of understanding this process.

This is not to say that we think the elimination of violent and controlling behaviours on the part of perpetrators should not be the ultimate goal of community responses; it most definitely should. We anticipate that one of the conclusions that may be drawn by this review is that significantly more work needs to be done to show how the achievement of these proximal outcomes ultimately can lead to the cessation of violence. However, as we look more closely at the literature surrounding BIPs, it has become apparent that factors unrelated to the programmes themselves also contribute to the likeliness that men will cease to engage in violence against partners and family members. Influences at both the individual and interpersonal levels, as well as the community and social levels, are at play, and these need to be identified and accounted for within the larger coordinated response.

REFERENCES

1. Gondolf EW. *Batterer intervention systems: issues, outcomes, and recommendations*. Thousand Oaks, CA: Sage, 2002.
2. Carter LS. *Batterer intervention: doing the work and measuring the progress*. A Report on the December 2009 Experts Roundtable. San Francisco: Family Violence Prevention Fund & National Institute of Justice, 2010.
3. Babcock JC, Green CE, Robie C. Does batterers' treatment work? A meta-analytic review of domestic violence treatment. *Clin Psychol Rev* 2004;23:1023–53.
4. Bowen E, Gilchrist E. Comprehensive evaluation: a holistic approach to evaluating domestic violence offender programmes. *Int J Offender Ther Comp Criminol* 2004;48:215–34.
5. MacLeod G, Pi R, Smith D, *et al*. Batterer intervention systems in California. California, Judicial Council of California/Administrative Office of the Courts, 2009.
6. Pepin D, Hoss L. *Menu of State batterer intervention program laws*. Atlanta, GA: Center for Disease Control and Prevention Office for State, Tribal, Local and Territorial Support, Public Health Law Program, 2015.
7. Maiuro RD, Eberle JA. State standards for domestic violence perpetrator treatment: current status, trends, and recommendations. *Violence Vict* 2008;23:133–55.
8. Miller M, Drake E, Nafziger M. *What works to reduce recidivism by domestic violence offenders*. Olympia, WA: Washington State Institute for Public Policy, 2013.
9. Dutton DG, Corvo K. Transforming a flawed policy: a call to revive psychology and science in domestic violence research and practice. *Aggression Violent Behav* 2006;11:457–83.
10. Gondolf EW. Theoretical and research support for the Duluth model: a reply to Dutton and Corvo. *Aggression Violent Behav* 2007;12:644–57.
11. Dixon L, Archer J, Graham-Kevan N. Perpetrator programmes for partner violence: are they based on ideology or evidence? *Legal Criminol Psychol* 2012;17:196–215.
12. Gelles RJ. The politics of research: the use, abuse, and misuse of social science data?The cases of intimate partner violence. *Fam Court Rev* 2007;45:42–51.
13. Sayer A. Critical realism and the limits to critical social science. *J Theory Soc Behav* 1997;27:473–88.
14. Pawson R. *Evidence-based policy*. London: SAGE Publications Ltd, 2006.
15. Jackson S, Feder L, Forde DR, *et al*. *Batterer intervention programs: where do we go from here*. Washington DC: U.S. Department of Justice, Office of Justice Programs, 2003.
16. Stover CS, Meadows AL, Kaufman J. Interventions for intimate partner violence: review and implications for evidence-based practice. *Prof Psychol Res Pract* 2009;40:223–33.
17. Wong G, Westhorp G, Pawson R, *et al*. Realist synthesis: RAMESES training materials: RAMESES Project. 2013.
18. Connelly JB. Evaluating complex public health interventions: theory, methods and scope of realist enquiry. *J Eval Clin Pract* 2007;13:935–41.
19. Pawson R, Greenhalgh T, Harvey G, *et al*. Realist review—a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;10(Suppl 1):21–34.
20. Jagosh J, Pluye P, Wong G, *et al*. Critical reflections on realist review: insights from customizing the methodology to the needs of participatory research assessment. *Res Synth Methods* 2014;5:131–41.
21. Sartin R, Hansen D, Huss M. Domestic violence treatment response and recidivism: a review and implications for the study of family violence. *Aggression Violent Behav* 2006;11:425–40.
22. Saunders D. Group interventions for men who batter: a summary of program descriptions and research. *Violence Vict* 2008;23:156–72.
23. Paymar M, Garnes G. *Countering confusion about the Duluth model*. Duluth, MN: Domestic Abuse Intervention Programs, 2007.
24. Price BJ, Rosenbaum A. Batterer intervention programs: a report from the field. *Violence Vict* 2009;24:757–70.
25. Mankowski ES, Haaken J, Silvergleid CS. Collateral damage: an analysis of the achievements and unintended consequences of batterer intervention programs and discourse. *J Fam Violence* 2002;17:167–84.
26. Gondolf EW. How batterer program participants avoid reassault. *Violence Against Women [HWWilson—SSA]* 2000;6:1204–222.
27. Rosenbaum A, Leisring PA. Group intervention programs for batterers. *J Aggression Maltreat Trauma* 2001;5:57–71.
28. Gilchrist E. The cognition of domestic abusers: explanations, evidence and treatment. In: Gannon T, Ward T, Beech A, *et al*, eds. *Aggressive offenders' cognition: theory, research, and practice*. John Wiley & Sons, 2007:247–66.
29. Holtzworth-Munroe A, Stuart GL. Typologies of Male batterers: three subtypes and the differences among them. *Psychol Bull* 1994;116:476–497.
30. Carlson RG, Jones KD. Continuum of conflict and control: a conceptualization of intimate partner violence typologies. *The Family Journal* 2010;18:248–54.
31. Gondolf EW. Who are those guys? Toward a behavioral typology of batterers. *Violence Vict* 1988;3:187–203.
32. Hamberger LK, Hastings JE. Personality correlates of men who abuse their partners: a cross-validation study. *J Fam Violence* 1986;1:323–41.
33. Johnson MP. *A typology of domestic violence: intimate terrorism, violent resistance, and situational couple violence*. Boston, MA: Northeastern University Press, 2008.
34. Buzawa ES, Buzawa CG, Stark E. *Responding to domestic violence: the integration of criminal justice and human services*. Thousand Oaks, CA: Sage, 2017.
35. Stuart GL, Temple JR, Moore TM. Improving batterer intervention programs through theory-based research. *JAMA* 2007;298:560–2.
36. Day A, Chung D, O'Leary P, *et al*. Programs for men who perpetrate domestic violence: an examination of the issues underlying the effectiveness of intervention programs. *J Fam Violence* 2009;24:203–12.
37. Saunders DG. Feminist-cognitive-behavioral and process-psychodynamic treatments for men who batter: interaction of abuser traits and treatment models. *Violence Vict* 1996;11:393–414.
38. Bennett L, Violence PCAD. *Controversies and recent studies of batterer intervention program effectiveness*. National Resource Center on Domestic Violence VAWnet, 2001.
39. Diefenbeck C. Group therapy for male batterers: comparison of cognitive-behavioral & object relations approaches. *J Psychosoc Nurs Ment Health Serv* 2003;41:18–25.
40. MacLeod D, Pi R, Smith D, *et al*. *Batterer intervention systems in California: an evaluation*. San Francisco, CA: Judicial Council of California, 2009.
41. Scott KL, Wolfe DA. Readiness to change as a predictor of outcome in batterer treatment. *J Consult Clin Psychol* 2003;71:879–89.
42. Melton H, Belknap J. He hits, she hits: assessing gender differences and similarities in officially reported intimate partner violence. *Crim Justice Behav* 2003;30:328–48.
43. Wong G, Greenhalgh T, Westhorp G, *et al*. RAMESES publication standards: realist syntheses. *BMC Med* 2013;11:21.
44. Gondolf EW. Evaluating batterer counseling programs: a difficult task showing some effects and implications. *Aggression Violent Behav* 2004;9:605–31.