

THE EVALUATION GUIDE:

A guide for evaluating behaviour change programs for men who use domestic and family violence

Angela Nicholas, Georgia Ovenden, Rodney Vlasis



ANROWS

ANROWS acknowledgement

This material was produced with funding from the Australian Government Department of Social Services. Australia's National Research Organisation for Women's Safety (ANROWS) gratefully acknowledges the financial and other support it has received from the government, without which this work would not have been possible. The findings and views reported in this paper are those of the authors and cannot be attributed to the Australian Government Department of Social Services.

Acknowledgement of Country

ANROWS acknowledges the Traditional Owners of the land across Australia on which we work and live. We pay our respects to Aboriginal and Torres Strait Islander Elders past, present, and future, and we value Aboriginal and Torres Strait Islander histories, cultures, and knowledge. We are committed to standing and working with Aboriginal and Torres Strait Islander peoples, honouring the truths set out in the [Warawarni-gu Guma Statement](#).

© ANROWS 2020

Published by

Australia's National Research Organisation for Women's Safety Limited (ANROWS)
PO Box Q389, Queen Victoria Building, NSW 1230 | www.anrows.org.au | Phone +61 2 8374 4000
ABN 67 162 349 171

Creative Commons Licence

Attribution-Non Commercial



CC BY-NC

This licence lets others distribute, remix and build upon the work, but only if it is for non-commercial purposes and they credit the original creator/s (and any other nominated parties). They do not have to license their Derivative Works on the same terms.

Version 3.0 (CC Australia ported licence): [View CC BY-NC Australia Licence Deed](#) | [View CC BY-NC 3.0 Australia Legal Code](#)
Version 4.0 (international licence): [View CC BY-NC 4.0 Licence Deed](#) | [View CC BY-NC 4.0 Legal Code](#)

Please note that there is the potential for minor revisions of this report.
Please check the online version at www.anrows.org.au for any amendment.

THE EVALUATION GUIDE:

A guide for evaluating behaviour change programs for men who use domestic and family violence

Dr Angela Nicholas

Centre for Mental Health, School of Population and Global Health, University of Melbourne

Dr Georgia Ovenden

Centre for Mental Health, School of Population and Global Health, University of Melbourne

Rodney Vlais

Consultant

This resource was developed as part of the ANROWS research project BW.19.01 “Development of a best practice guide to perpetrator program evaluation”. Please consult the ANROWS website for more information on this project and the full research report: Nicholas, A., Ovenden, G., & Vlais, R. (2020). *Developing a practical evaluation guide for behaviour change programs involving perpetrators of domestic and family violence* (Research report, 17/2020). Sydney: ANROWS.

Suggested citation

Nicholas, A., Ovenden, G., & Vlais, R. (2020). *The Evaluation guide: A guide for evaluating behaviour change programs for men who use domestic and family violence* (ANROWS Insights, 02/2020). Sydney: ANROWS.

Contents

List of figures	3
List of tables	4
List of boxes	4

Introduction 5

Purpose of this Evaluation guide	5
Australian context	5
Structure of this guide	6
What is evaluation?	6
Why evaluate?	6
Before you begin: Evaluation readiness	7

Scoping the evaluation 9

Understanding the program's theoretical framework	9
Articulating the program goal	9
Developing a program logic:	
Describing the program and its mechanisms	11
Developing and prioritising evaluation questions	17
Other considerations regarding measuring outcomes	23

Answering the evaluation questions 27

Choosing key indicators to answer your evaluation questions	27
Choosing data collection methods to answer the evaluation questions	31
Selecting an appropriate and achievable study design	36
Deciding who should conduct your evaluation	44
An overarching consideration:	
Partnerships between program providers and evaluators	44

Ethical considerations	46
Principles of ethical research	46
Ethical guidelines specifically for evaluation	46
Management and collection of data from victims/survivors	49
Ethical approval and working with vulnerable groups	51
Further reading	52
References	55

List of figures

FIGURE 1:	Components of the program logic model	13
FIGURE 2:	A hypothetical simple logic diagram for an MBCP adapted from a program logic by No to Violence	15
FIGURE 3:	A hypothetical “if ... then” program logic for an MBCP adapted from a program logic by No to Violence	16
FIGURE 4:	Steps in developing evaluation questions	17
FIGURE 5 :	Examples of evaluation questions that can be linked to the program logic in Figure 3	19
FIGURE 6:	Indicators to answer evaluation questions arising from the program logic in Figure 3	28
FIGURE 7:	Possible data collection methods used for indicators outlined in Figure 6	34

List of tables

TABLE 1:	Examples of process evaluation questions arising from the program logic in Figure 3	20
TABLE 2:	Examples of outcomes evaluation questions devised from the program logic in Figure 3	22
TABLE 3:	Examples of questions and indicators for assessing unintended consequences of MBCPs	26
TABLE 4:	How to assess indicators against the SMART criteria	29
TABLE 5:	Advantages and disadvantages of various data collection methods	32
TABLE 6:	Examples of types of study design	38
TABLE 7:	Advantages and disadvantages of internal and external evaluators	45

List of boxes

BOX 1:	A note on “outputs”	13
BOX 2:	Process evaluation example	21
BOX 3:	Assessing transferability of MBCPs	26
BOX 4:	Non-equivalent control group design: Domestic violence perpetrator programs—Steps towards change (Kelly & Westmarland, 2015)	40
BOX 5:	Longitudinal evaluation: European model for MBPC evaluations	41
BOX 6:	Challenges of longitudinal evaluation: A study of the impact on men and their partners in the short term and long term of attending MBPCs (Brown et al., 2016)	42

THE EVALUATION GUIDE:

A guide for evaluating behaviour change programs for men who use domestic and family violence

Introduction

Purpose of this Evaluation guide

The purpose of this Evaluation guide is to provide easy-to-understand information on scoping an evaluation for men's behaviour change programs (MBCPs) focusing on men who use domestic and family violence (DFV). This Evaluation guide aims to improve the technical knowledge for personnel involved in the implementation of MBCPs who might be involved in commissioning an external evaluation or conducting an in-house evaluation. Upon working through this Evaluation guide, readers should:

- understand the purposes of conducting quality evaluations of MBCPs
- be able to develop a program logic and articulate appropriate evaluation questions
- be aware of a range of methodologies available to answer the evaluation questions
- gain an awareness of the important considerations in designing an evaluation of MBCPs
- be aware of the ethical issues that need to be considered when commissioning or conducting an evaluation of an MBCP.

The information contained in the guide will contain excerpts from interviews (presented in some of the boxes) with evaluators of MBCPs across Australia and a group consultation conducted in Melbourne, Victoria, involving program managers and facilitators of MBCPs.

Australian context

There are numerous definitions of MBCPs in use in Australia. For the purpose of this guide, we have limited our discussion to group work programs involving men who use violence and controlling behaviour against women. However, the evaluation principles outlined in this guide may be applicable to other perpetrator interventions, and indeed all DFV behaviour change programs. We also recognise that not all victims/survivors of DFV are women, nor all perpetrators men; however, for the purposes of this study, we use the term DFV to refer to that perpetrated against women by men.

This guide has also been informed by a state of knowledge review that assessed the current state of practice standards for MBCPs in place in Australia. This review also

examined the academic and grey literature on outcome measures used in MBCPs in the domains of:

1. long-term changes in perpetrators' violent and controlling behaviour
 2. adult victim/survivor safety, wellbeing and freedom
 3. children's safety, wellbeing and family functioning.
- (Adapted from Respect, 2017)

We also recognise that MBCPs are part of a complex environment and service system, and that an MBCP can only be as effective as the broader system within which it operates. The broader system elements may influence the success of the program and should be identified and considered in the design of any MBCP and its evaluation.

Structure of this guide

Following this introduction, the guide is divided into sections on scoping the evaluation and ethical issues associated with conducting an evaluation of an MBCP. The Scoping the evaluation section is further divided into six main sub-sections:

1. understanding the program's theoretical framework
2. articulating the program goal
3. developing a program logic
4. developing and prioritising evaluation questions
5. answering evaluation questions
6. deciding who should conduct the evaluation.



What is evaluation?

Formal program evaluation is “the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming” (Patton, 1997, p. 23).

Why evaluate?

Evaluation of MBCPs for perpetrators of DFV allows us to ask important questions that can help identify points in the implementation of the program where improvements can be made, as well as identify if the program is achieving its aims. Based on these evaluation findings, important decisions are often made, such as whether to continue, expand, modify or discontinue the program. In the context of an innovative or new type of program, evaluation is essential to help funders decide whether the innovation should be scaled up.

The other important purpose of evaluation is to contribute to the evidence base on MBCPs; that is, to contribute to the general knowledge in the sector such that others can use the knowledge gained to inform their own program. This is especially important in the area of MBCPs, where the evidence base is limited. There is always

a need to balance the responsibility of contributing to the evidence base with accountability for understanding the impact of an MBCP program.

Before you begin: Evaluation readiness

This guide takes you through the key steps of scoping an evaluation before you commence your own evaluation activities or commission someone to do this on your behalf. In this way, you are making sure that your MBCP is “evaluation-ready”. Determining whether your MBCP is evaluation-ready helps you to decide whether, at that certain point in time, an evaluation is “justified, feasible, and likely to provide useful information” (Kaufman-Levy & Poulin, 2003, p. 4). Evaluation readiness is also often called “evaluability” (Hawe, Degeling, & Hall, 1990). Completing an evaluation before it is evaluation-ready means that the outcomes measured may not accurately reflect the true effects of the program, but rather that the MBCP is not yet being fully implemented.

Day et al. (2019), in their recent Australia’s National Research Organisation for Women’s Safety (ANROWS) research report on improving the quality of MBCPs, argued that one of the major factors hampering the development of a knowledge base in this field is the lack of evaluation readiness of programs being formally evaluated. The authors pointed, in particular, to the lack of guiding program logics, lack of clarity of theoretical frameworks underpinning programs, and program integrity in relation to evaluated programs. Program integrity refers to the degree to which the program is being implemented as intended (also sometimes called “fidelity”).

Completing the steps in this Evaluation guide in order to scope the evaluation will assist you in deciding whether the program is evaluation-ready. However, there are several other questions you can consider when attempting to gauge whether a program has sufficient program integrity to be evaluation-ready. These include the following:

- Are program staff, including group work facilitators, other practitioners and partner contact workers, sufficiently trained in MBCP work in general, as well as in the particular theoretical orientation or approach adopted by the program?
- Are internal and external supervisors aware of, and able to talk about, the theoretical approaches and behaviour change models adopted by the program, and can they identify when practice is drifting in unintended directions?
- Can live or recorded practice be observed to enable reflection on the moment-by-moment ways in which practitioners attempt to implement the program’s theoretical approaches and behaviour change models?
- Does program management, and the agency’s ethos, support the theoretical orientation taken in the program?
- Are the program’s theoretical underpinnings, conceptual foundations and assumptions documented in a way that program practitioners can understand and follow, and in a way that assists them to translate these underpinnings and foundations into practice?
- Does the program guide or operational manual have sufficient detail to guide facilitation and practice, yet also retain the flexibility to enable practitioners to be responsive rather than prescribing rigidly controlled practice? (Vlais et al., 2017)

Monitoring program integrity involves more than determining if interventions with men perpetrating DFV are consistent with the program's theoretical underpinnings. It should also involve gauging whether the program's risk identification, risk assessment and risk management procedures are operating according to agency policy and jurisdiction-wide MBCP minimum standards, and whether the program is operating as part of an integrated response system.

It is also important to determine whether partner and family safety contact with victims/survivors is being implemented as planned and according to minimum standards. A program that acts with fidelity according to its model of behaviour change but does not put into practice intended policies and procedures regarding risk management and victim/survivor support is not evaluation-ready.

Many of the above considerations are process evaluation questions and might not be known until the evaluation is underway. When there is significant doubt about whether a program is acting with fidelity according to its theoretical underpinnings and conceptual approach, it can be beneficial to split the evaluation in two halves—first, a process evaluation to determine how program implementation might need to be strengthened and, after process evaluation adjustments are made, a second evaluation stage focusing on measuring outcomes.

More information on evaluation readiness and related tools is available from many sources, including:

- *Evaluability assessment: Examining the readiness of a program for evaluation.* <http://www.jrsa.org/pubs/juv-justice/evaluability-assessment.pdf> (Kaufman-Levy & Poulin, 2003).
- Better evaluation: Evaluability assessment. https://www.betterevaluation.org/en/themes/evaluability_assessment (Davies, 2015).



Scoping the evaluation

Understanding the program’s theoretical framework

Most MBCPs are grounded in a theory of how they will work to achieve desired outcomes, and borrow from a range of theoretical viewpoints about *why* domestic and family violence occurs and *how* it might be stopped (Mackay et al., 2015; Paymar & Barnes, 2007). Understanding the framework that underpins your MBCP is important because it is this framework that will inform the program logic. The theoretical approach of the MBCP was likely adopted in the early stages of planning and program development and will have informed the content, structure and design of the program.

Theoretical frameworks relate not only to the mechanisms by which a program aims to facilitate changes in men’s behaviour. They also relate to other mechanisms through which the program attempts to work towards desired outcomes, such as frameworks informing partner contact and support for victims/survivors, and those informing the program’s approach towards coordinated and collaborative practice with other agencies. Multiple frameworks can therefore inform different components of the program.

Articulating the program goal

A clear understanding of the MBCP’s overarching goal ensures that program staff, evaluators and other stakeholders have a shared understanding of what the program is working towards. The goal is a broad statement of what the MBCP aims to achieve in the long term: your program’s “mission”.

There is an emerging consensus in the perpetrator intervention literature that all perpetrator interventions should aim to achieve safety and wellbeing for women, children and others who experience men’s use of DFV. While program goal statements need to be relatively brief, it is important for the program and its evaluators to be clear about the meaning of terms used in the goal statement. For example, how is *safety and wellbeing* defined by the program? Does the term incorporate emotional and other forms of safety, in addition to physical safety? Does *wellbeing* include the strength of relationships between adult and child victims/survivors in the family that might have been harmed by the perpetrator’s actions? The conceptualisation of key terms should be articulated in program and evaluation documentation following the program goal statement.

Presently, the most recognised understanding of “success” in MBCPs is typified by that outlined in the Project Mirabal study (Kelly & Westmarland, 2015). This United Kingdom-based project measured program success as “the extent to which perpetrator programs reduce violence and increase safety for women and children, and the routes by which they contribute to coordinated community responses to domestic violence” (Kelly & Westmarland, 2015, p. 7).

Project Mirabal (Kelly & Westmarland, 2015) included interviews with female partners and ex-partners, male participants, practitioners and funders to identify what success

in their MBCP meant to them. The results revealed that women were focused on six outcomes of success, which moved beyond stopping the violence and included the following:

1. respectful communication
2. expanded space for action for women, which restores their voice and freedom to make choices, while improving their wellbeing
3. safety and freedom from violence
4. safe, positive and shared parenting
5. enhanced perpetrator awareness about the impact of violence on others
6. safer lives for their children. (Kelly & Westmarland, 2015)

It is likely that your program goal will include some or all of these outcomes. However, programs may have other goals that are worth articulating. For example, the aim of the Caring Dads group intervention program is

to engage fathers who have used violence, to help them develop skills in child-centred fathering and take responsibility for the impacts of their violence upon their children and their children's mother. (Diemer et al., 2020, p. 12)



More information on articulating program goals can be found at https://www.betterevaluation.org/en/rainbow_framework/frame/decide_purpose (Better Evaluation, 2016a).

Quality of life indicators

Recent Australian research has focused on proposing a set of women's quality of life (QOL) indicators as outcomes for evaluating MBCPs (McLaren, Fischer, & Zannettino, 2020). One hundred women, 71 of whom had partners who had participated in a DFV perpetrator intervention program, were asked what quality of life meant to them. The most frequently endorsed QOL themes were autonomy, informal supports (family and friends), emotional health, safety (physical and psychological), children and pets, and mental health.

The authors argued that measuring women's QOL indicators before and after their (ex-)partner's participation in an MBCP provides a way of determining whether women's lives have improved, without needing to focus research interviews directly on the women's experiences of DFV (McLaren et al., 2020). They speculated that this might have potential in increasing women's participation rates in MBCP outcome evaluation studies, as the women would not be required (unless they choose) to talk directly about the violence or about the man's participation in the program. Given the substantial impact of DFV across a range of facets of women's lives, employing QOL indicators offers a promising avenue to focus outcome measures on the fundamental goal of MBCPs.

Developing a program logic: Describing the program and its mechanisms

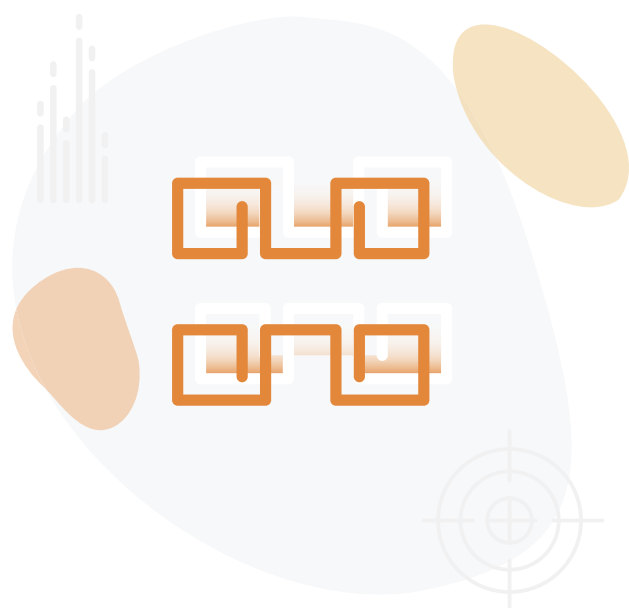
A program logic is a systematic way to present and share your understanding of the relationships between the resources used in the MBCP program, the activities you plan, and the changes or results you hope to achieve (Kellogg, 2004), including your articulated program goal. A program logic, if sufficiently detailed, articulates the “mechanisms” of how each of the MBCP activities leads to men reducing and ultimately ending their violent and controlling behaviour and to victims/survivors experiencing greater safety and wellbeing.

Why do you need a program logic?

Articulating the theoretical mechanism is important for many reasons:

- It can help you and all practitioners involved in running the program understand how the activities of the program are expected to lead to the desired outcomes.
- It is a useful device to use in collaboration with stakeholders to create a shared understanding of the desired outcomes of an MBCP program and the activities designed to achieve these (McKenzie, White, Minty, & Clancy, 2016).
- It can identify possible barriers to the quality implementation of the activities and the achievement of the desired outcomes (i.e. it helps to identify points at which the program implementation might “fall down” or where preceding outcomes may not lead to subsequent ones).
- It helps to inform the evaluation, particularly by helping in the selection of evaluation questions.

Logic models are not necessarily static, and can be revised or updated with continued program learning through formal and informal evaluation.



EXAMPLE**Developing the program logic with evaluation users**

For one evaluator we interviewed, the development and consideration of a program logic model had a significant impact on the evaluation users and helped to crystallise the evaluation questions:

Participants found the program logic workshops very useful, as it enabled them to reflect on the fullness of the intent of the programs, to consider differences between sites, and for the facilitators to consolidate their thoughts regarding their expectations and hopes for the program. As well as being useful for the program teams, it also helped the evaluation team to develop questions for future interviews with the facilitators.

Another evaluator suggested that adequate involvement of practitioners in the development of the program logic was essential to improve the external evaluators' understanding of the program:

Evaluators need to work collaboratively with the practitioners to develop the evaluation [program] logic; this helps practitioners to be on side with the evaluation, because the evaluation then reflects how the practitioners perceive the program and what's important for them to find out through the evaluation. Evaluators can come in with their own narratives or stories regarding the priorities for the evaluation, shaped by a range of influences (including expectations from the commissioner of the evaluation), and if there isn't sufficient involvement of practitioners in the development of the program logic, the essence of what needs to be evaluated can become lost.

The importance of collaboration was raised by another evaluator, who underlined the essential role of evaluators "doing up-front work" and developing strong working relationships with the MBCP practitioners. Co-designing the program logic might be one way to build this relationship:

Time and relationship-building effort is required for the program practitioners to be on board with the evaluation. Them being on board is crucial for things like recruiting program participants to be interviewed. If they aren't on board, evaluators can encounter a whole lot of problems across the evaluation. It's worth investing the collaboration-building time.

Developing the program logic with evaluation users

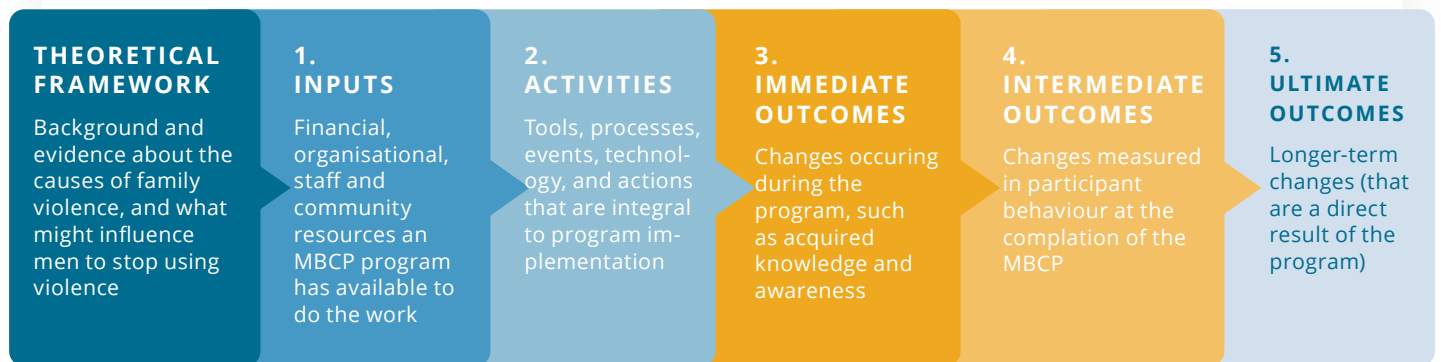
Where possible, it can be useful to develop the program logic with the evaluation users. The involvement of users in the development of the logic could range from providing feedback on a first draft written by the evaluator, to collaborative workshops where the logic is "co-designed". While program funders, especially those commissioning the evaluation, have an important stake in collaborating in program logic design, the program providers themselves will be in the strongest position to co-design the logic with the evaluator. Ideally, the program provider would have already constructed a program logic as part of the program's initial development process, but this is not always the case. The examples below show some of the advantages of this collaborative approach to the development of the program logic for the evaluators and for the evaluation users.

Key components of the program logic

The key components of a program logic are illustrated in Figure 1. It is worth noting here that there is much variation in the language used in evaluation, with different sources calling the same concepts by different names. For example, what we have

called “outcomes” are sometimes called “impacts”, and “inputs” are sometimes called “resources”. For simplicity and consistency, we use the terms outlined in Figure 1 throughout this guide. “Outputs” might be included in your program logic (see Box 1).

FIGURE 1:
Components of the program logic model



BOX 1:
A note on “outputs”

Outputs relate to the “products” of the activities of the MBCP. For example, an output of an MBCP risk review meeting (the activity) could be risk management plans for higher risk men (the output). Outputs are often confused with the indicators of the activities—for example, “number of MBCP group work sessions” as an output measure of the activity “MBCP group work sessions delivered”. Therefore, they often duplicate the activity, sitting alongside it in the program logic. The program logic should aim to demonstrate the hypothetical mechanisms of the program or how the program is supposed to work. The number of MBCP group work sessions is instead one indicator of how this activity was implemented. Outputs can be included in program logics, but take care that they are not just replicating the activity as many logic diagrams do.

Adapted from Knowlton and Phillips (2012)

The framework represented in Figure 2 is used to present a simple program logic for a hypothetical MBCP program. While this program logic illustrates the inputs, general program activities and immediate, intermediate and ultimate outcomes, the linkages between each are minimal. The program logic presented in Figure 3, however, is an example where a greater number of linkages can be used to illustrate the theoretical mechanism underpinning the program in more detail. Detail such as this can inform program design and identify potential weak points on which to focus additional resources, or where unintended consequences of the program might arise. If we look at Figure 3 under “activities”, at the first box: if the perpetrators are not deemed “ready” for the group-based component of the program as determined by the intake assessment criteria for suitability, but are nevertheless immediately enrolled in the group work component of the program without one-to-one sessions to improve readiness, then the program is unlikely to achieve some of its intended outcomes. In fact, we might instead see negative unintended outcomes as a result, whereby a perpetrator who was not ready at time of intake might undermine the facilitator in the group work or use some of the information provided to expand his controlling behaviours.

Focusing on multiple change mechanisms in the program logic

It is important that your program logic captures all of the main mechanisms and components through which your program works towards its ultimate goal. MBCPs work towards desired ultimate outcomes through multiple mechanisms. Attempting to change men's violence-supporting attitudes and behaviours is, of course, one of them. However, other mechanisms include the provision of partner and family safety contact, and risk management processes that identify and respond to perpetrator-driven risk. For MBCPs to operate safely, ethically and according to jurisdiction-based minimum standards, mechanisms to provide victims/survivors with partner and family safety contact, to identify and respond to perpetrator-driven risk, and to collaborate with other agencies as part of an integrated system response, are as important as mechanisms that work towards changing perpetrator attitudes and behaviours.

By focusing only on the mechanisms through which the program works towards men's attitudinal and behavioural change, the impacts of other mechanisms that work towards the ultimate program goal remain invisible. The evaluation then captures only part of the story.

FIGURE 2:

A hypothetical simple logic diagram for an MBCP adapted from a program logic by No to Violence

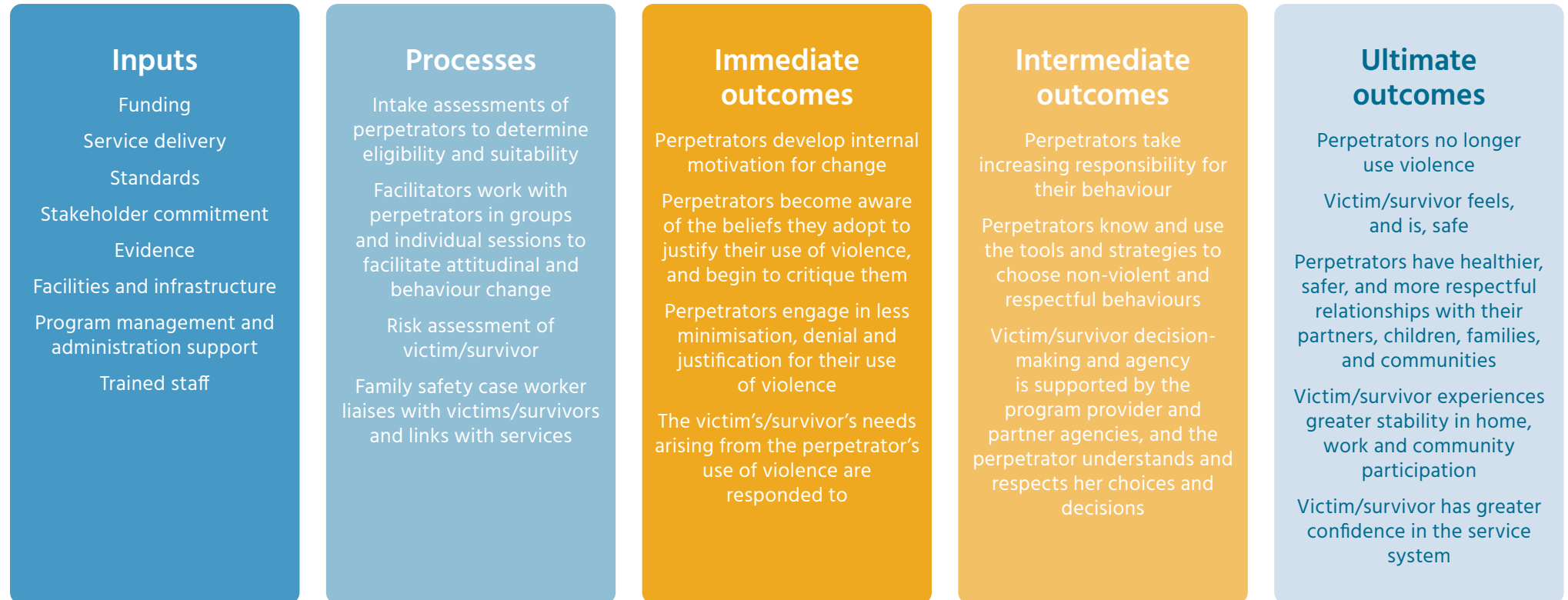
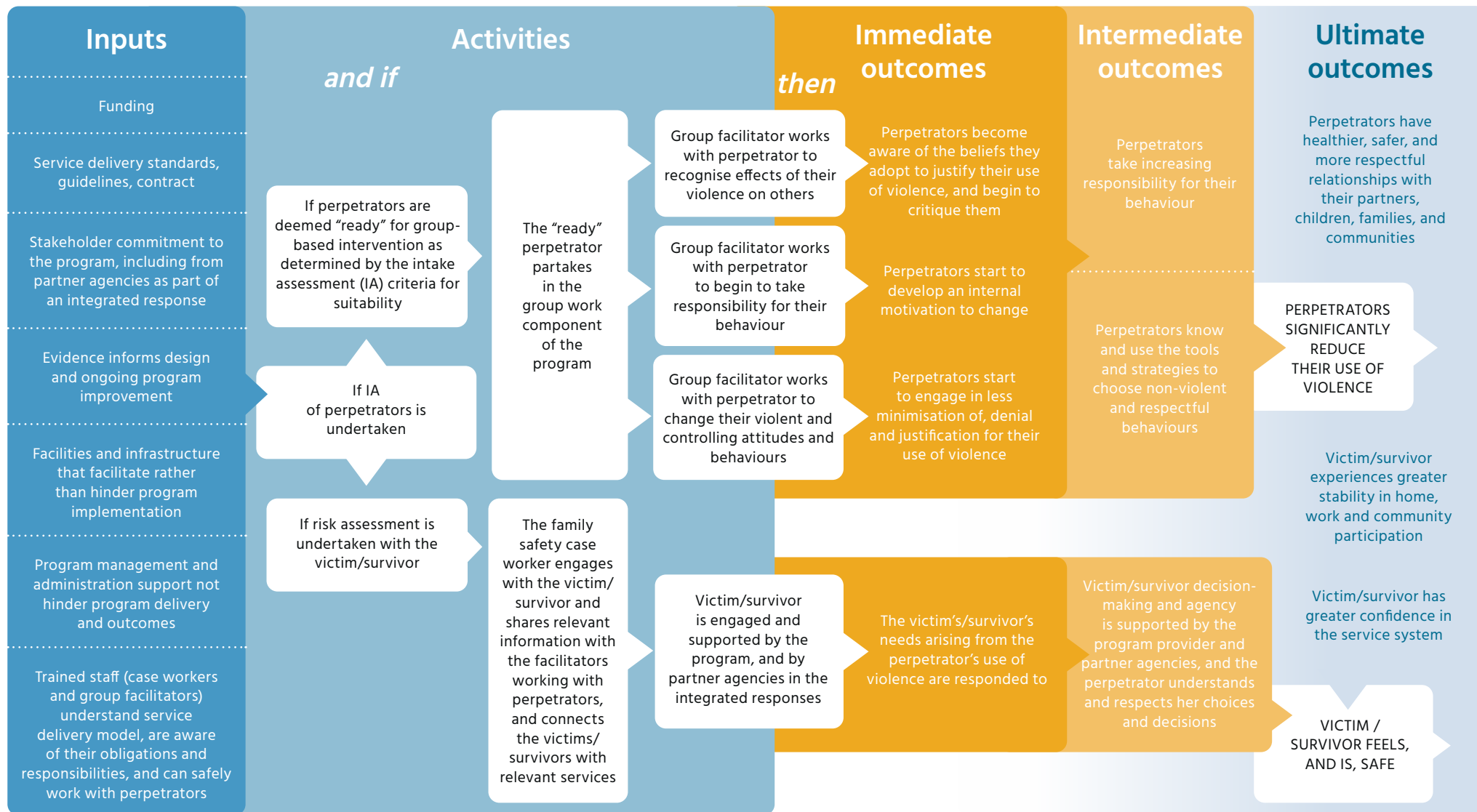


FIGURE 3:

A hypothetical “if ... then” program logic for an MBCP adapted from a program logic by No to Violence



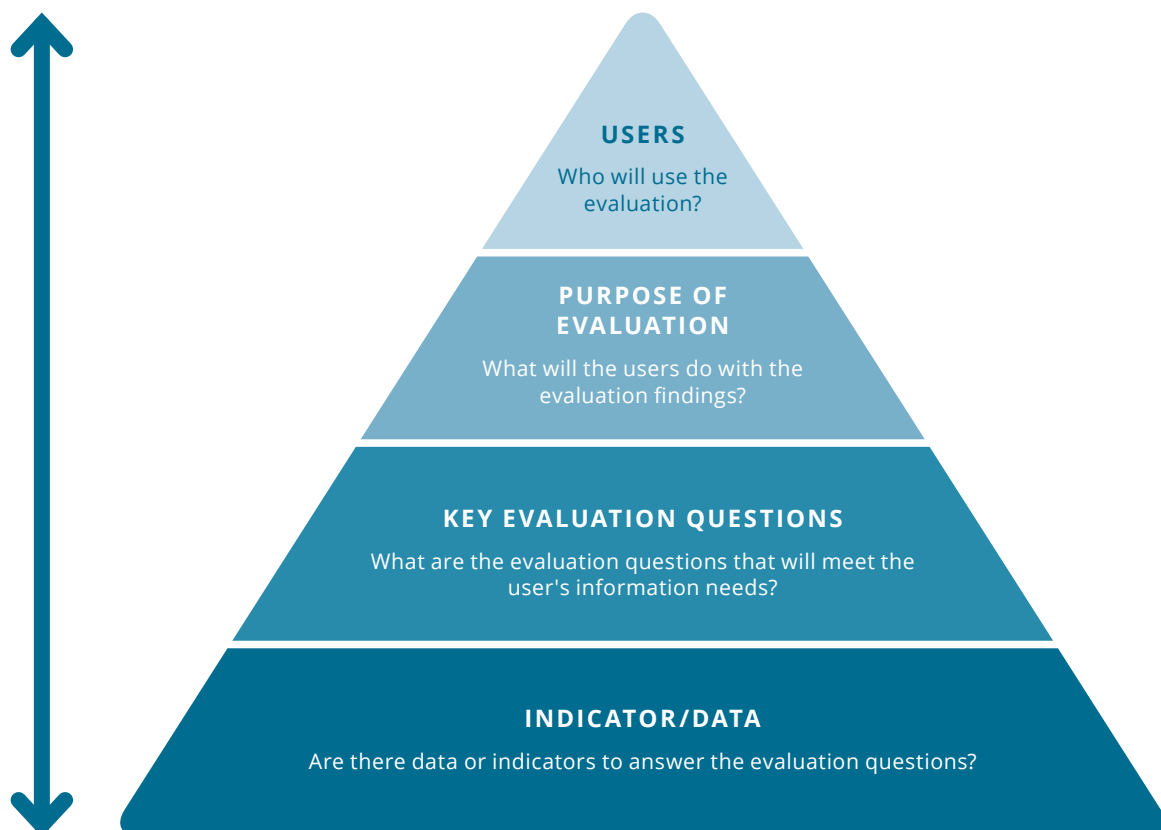
Developing and prioritising evaluation questions

Evaluation questions can relate to any point in the program logic: the inputs, for example, “Is there support for the program in this community?”; the activities, for example, “Are victims/survivors engaged and supported by the program?”; and the outcomes, for example “Does the MBCP improve the safety of women and children?” (ultimate outcome). There are many possible evaluation questions, so how do you decide which to choose? In this section, we consider the steps to creating useful key evaluation questions. Figure 4 illustrates that the evaluation users’ information needs should inform the purpose of the evaluation, which should in turn determine the evaluation questions. From there, the evaluation questions should inform the indicators or data collected. The arrows illustrate the linkages between the users’ needs, the purpose of the evaluation, the key questions and data.



More information on developing evaluation questions can be found at https://www.betterevaluation.org/en/rainbow_framework/frame/specify_key_evaluation_questions (Better Evaluation, 2016b).

FIGURE 4:
Steps in developing evaluation questions



Working with evaluation “users” (stakeholders)

It is wise to create evaluation questions in consultation with key stakeholders who are likely to be “users” of the evaluation (Figure 4). Evaluation users are people or groups who will or could use evaluation findings to make decisions about the program. Different users will have varying views on the most important type of information to gather from the evaluation because they will use that information for different purposes (Figure 4). Users might be people who have inputs into the program (e.g. government funders), whose actions are affected by the program (e.g. victims/survivors), and/or those who interact directly with the program (e.g. participants, program managers, facilitators and referrers). When working with stakeholders to identify evaluation questions, it is important to consider “who will use the evaluation and for what purpose”. Therefore, it is a useful step to list your evaluation users and consider consulting with each of them regarding how they want to use the results of the evaluation.

Engaging users in developing the evaluation questions will increase the usefulness, relevance and credibility of the evaluation. Thinking strategically, users who are engaged in the evaluation process are also more likely to support it and act on any outcomes or recommendations that arise from the evaluation.



More information on identifying the evaluation users and their uses for the evaluation can be found here: <https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/47278/133624.pdf?sequence=1&isAllowed=y> (International Development Research Centre, 2012).

Determining which evaluation questions to include

There will be many possible evaluation questions, especially when there are multiple evaluation users, each with their own priorities. As one of the participants in our consultation stated:

There is such a diversity in the cultures of stakeholders, both in terms of practice and also in terms of what they want from evaluation ... there is often clashing with stakeholders, who might want very different things.

Once you understand how evaluation findings will be used you may need to prioritise some evaluation questions, depending on a number of factors such as:

- “non-negotiable” questions required by the funder
- resources available (e.g. budget)
- timing of the evaluation (how long you have to do it)
- decisions to be informed by the evaluation (determined by consulting with users)
- which stakeholder groups to be consulted.

For instance, it might be a condition of the program funding agreement to evaluate and report on particular aspects of the MBCP. For some stakeholder groups (e.g. funders and program managers), it would be useful to determine decisions expected to be informed by the evaluation (e.g. funding decisions and program resourcing) and when they will be made to ensure that evaluation findings are relevant and timely. There also might be an opportunity for presenting the findings (e.g. a national conference, community forum, or organisational strategic review).

Negotiation is sometimes required to ensure that evaluation questions are realistic for budget, timeframe and data collection. Some stakeholders might be most interested in questions related to ultimate outcomes, which in the MBCP field often require relatively large evaluation budgets and an evaluation timeframe upwards of 2 years. Some stakeholders might place less value on process evaluation and whether an MBCP is being implemented with integrity.

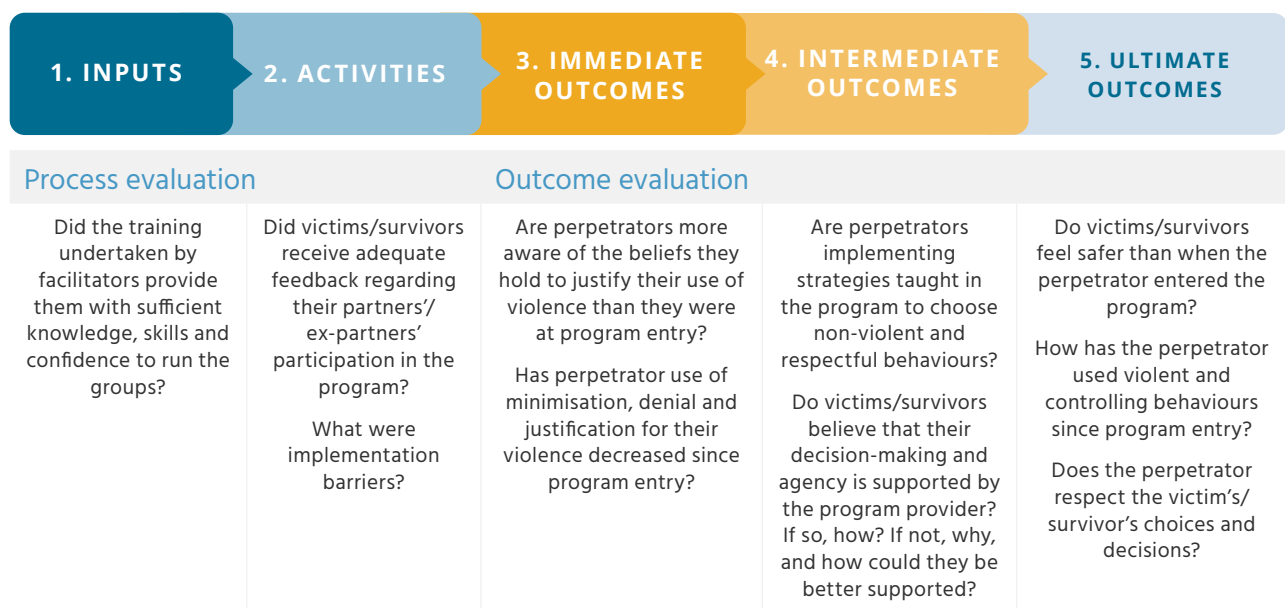
The program logic will also assist in deciding the order of answering evaluation questions. Evaluation questions on inputs and processes, for example, may need to be answered before questions about outcomes. This also relates back to evaluation readiness, where it is important to know if your program is being implemented properly before you explore program outcomes.

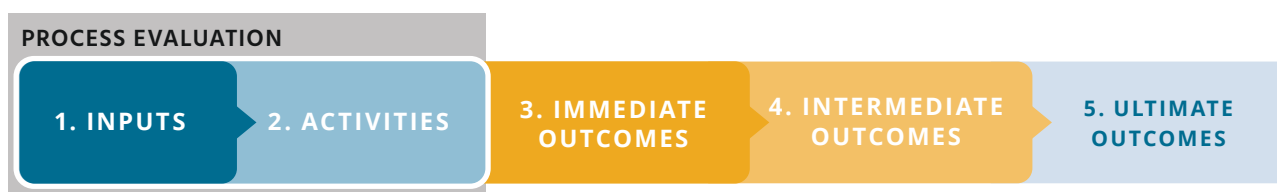
Developing evaluation questions from your program logic

Generally speaking, there are two main types of evaluation questions: process evaluation (how the program was delivered) and outcome evaluation (MBCP achievements). Figure 5 gives an example of how to use the program logic illustrated in Figure 3 to map evaluation questions.

FIGURE 5:

Examples of evaluation questions that can be linked to the program logic in Figure 3





Process evaluation

Process evaluation involves describing and assessing program inputs and activities and linking these to outcomes. Process evaluation questions are primarily concerned with the implementation, or process, of delivering MBCP activities. Process evaluation aims to assess how well the program is working, whether it is being implemented as designed (sometimes called “fidelity”), and whether it is accessible and acceptable to the target population. Process evaluation provides data and information about which aspects of the MBCP might need to be addressed or improved.

MBCP evaluators we interviewed placed a strong emphasis on process evaluation and included questions about participant recruitment (referral processes, participant characteristics, program operation), retention (in the overall service system), facilitator training and issues with program integrity. Importantly, process evaluation is key to determining evaluation readiness and whether the program has been implemented in a way that enables the MBCP to be ready for an outcome evaluation.

Process evaluation can reveal the internal workings of the program, but also how the system in which the program sits affects its implementation—for example, how men are being referred into the program, whether they are sufficiently prepared by the referrer and come into the program with realistic expectations, the overall impact of these factors on program retention, and how the program provider and other agencies work together to share information and manage risk. In this way, process evaluation helps determine whether the program has been implemented as planned, sits appropriately within the broader system, and is ready for an impact and outcome evaluation.

Process evaluation questions should integrate components of the program logic (i.e. inputs and activities). Table 2 shows examples of this.

TABLE 1:

Examples of process evaluation questions arising from the program logic in Figure 3

Process evaluation questions	Indicative questions
Process evaluation questions relating to inputs	<ul style="list-style-type: none"> • Is the program in line with current evidence on what works in MBCPs? • What is the cost per participant of running the program? Is it being sufficiently funded? • How can program managers better assist facilitators to implement the program?
Process evaluation questions relating to activities	<ul style="list-style-type: none"> • Does the current intake process adequately identify perpetrators who are suitable for the MBCP? • Is there sufficient communication between the victim/survivor case worker and the facilitator to inform the facilitator’s work with perpetrators in the program? • Are participants satisfied with MBCP activities? • Does the victim/survivor whose partner/ex-partner is involved in the MBCP feel their needs have been met? • Is the program encouraging perpetrators to take responsibility for their behaviour? If not, why not? • Which aspects of the facilitated session delivery can be improved and how? • What are program implementation barriers?

BOX 2:

Process evaluation example

An important question: “What are the barriers to implementation of the program?”

One MBCP evaluator reported examples of facilitators’ answers when asked about barriers to implementation. Some barriers were systems-generated, such as not receiving enough referrals to the program or a lack of clarity about program processes such as how to report back to referrers. Barriers were assessed at numerous time points to explore how facilitators addressed barriers throughout MBCP implementation.

A systems-level approach to process evaluation

Evaluators explained the importance of a systems-level approach to process evaluation:

New MBCPs can struggle to get up and going because they rely on establishing and strengthening new referral pathways with other agencies and sectors. When a new program arises, potential referrers might misconstrue what the program is about, have unrealistic expectations concerning what it can achieve, it might be difficult to recruit suitable trained practitioners who practice in a way that’s consistent with the program’s theoretical approach, et cetera. A process evaluation enables these and other issues to be explored, rather than wasting evaluation dollars at such early stages on evaluating impact and outcome—this can be a waste, as if the program hasn’t yet got off the ground in the way that it’s intended, then an impact evaluation is premature.

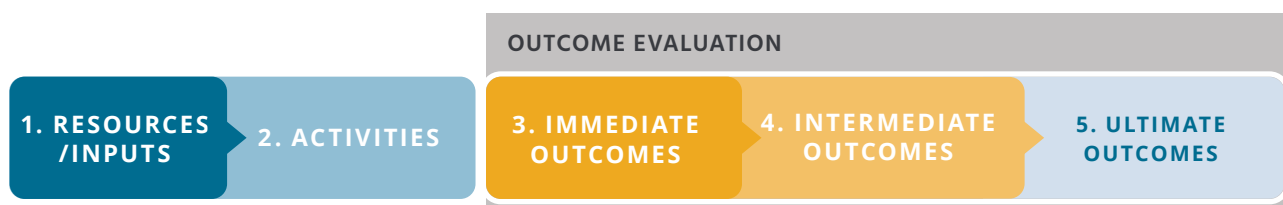
The extent to which a program works is dependent heavily on the system in which the program is embedded. The evaluation’s methodology was finely tuned to focus on this system as this MBCP represented a new program in Australia. What was most pressing in this evaluation was how this new program could be introduced successfully into a pre-existing system. How do other services relate to it? The focused process evaluation methodology, including interviews with referrers, and the systems focus, enabled them to get at the roots of any barriers, and what would be needed to address these barriers.

An overarching consideration:**Working with program providers throughout the evaluation**

One evaluator raised the ethical issue in evaluation design of reporting back to program providers throughout the evaluation, saying that it was unethical when conducting a long-term evaluation to wait until the end to report on findings. This issue is particularly relevant to MBCP evaluations when interim findings indicate that program activities seem to be having a positive or negative impact on victim/survivor safety, suggesting that certain activities should be either scaled up or changed.

Having an ongoing relationship with providers is also important for evaluation quality. A close relationship can identify difficulties early and ensure they are resolved such that they do not threaten the integrity of the evaluation. For example, it would be important to identify if service providers do not understand what data needs to be collected or how to record it and then provide clarification and assistance to ensure data integrity.

If changes to the program are made during the evaluation on the basis of interim or ongoing evaluation results, these need to be documented in the evaluation report and considered in an evaluation of program outcomes.



Outcome evaluation

Outcome evaluation focuses on the immediate, short- and long-term effects of an MBCP on its target populations (i.e. perpetrators, women and children). Generally, outcome evaluations of MBCP programs aim to provide evidence of change in attitudes, motivations, awareness, behaviours and experiences (e.g. victims'/survivors' safety and wellbeing) over time. Outcome evaluation questions should reflect the outcomes depicted in your logic model (i.e. immediate, intermediate, and ultimate outcomes). Some examples are included in Table 3.

TABLE 2:

Examples of outcomes evaluation questions devised from the program logic in Figure 3

Outcome evaluation questions	Indicative questions
Immediate outcomes evaluation questions	<ul style="list-style-type: none"> • Are perpetrators challenging beliefs justifying their use of violence upon completion of the program? • Are perpetrators engaging in less denial and justification of their use of violence than when they entered the program? • Are victim/survivor needs arising from the perpetrator's use of violence being responded to?
Intermediate outcomes evaluation questions	<ul style="list-style-type: none"> • Are perpetrators demonstrating increased use of non-violent and respectful behaviours toward their partner from when they entered the program? • Are perpetrators respecting the choices and decisions of their partner/ex-partner more than when they entered the program? • Do victims/survivors believe that their decision-making and agency is being supported by the program provider and partner agencies?
Ultimate outcomes evaluation questions	<ul style="list-style-type: none"> • Are perpetrators who have completed the program engaging in fewer violent, coercive and controlling behaviours and is this change sustained over time? • Are victims/survivors experiencing greater stability in their life (home, work, community participation) than when the perpetrators entered the program? • Are perpetrators relating to children in more respectful and child-centred ways than when they entered the program?



Evaluating MBCP ultimate outcomes

The overarching goal of MBCP evaluation is to examine if the outcomes of increased safety and wellbeing for women and children are achieved, and if these outcomes are sustained over time (Gondolf, 2012). It is important to note that the outcome of “increased safety and wellbeing for women and children” may be conceptualised in numerous ways. For example, is it about preventing injury from physical violence? Is it about victims/survivors’ perceived feelings of safety? Is it about freedom to live their lives and release them from their partner’s coercive control? Or is it about improvements in quality of life indicators that matter most to women and children?

It is important that ultimate outcomes measures focusing on victim/survivor safety do so in ways that are not solely linked to changes in the perpetrator’s behaviour. Victim/survivor safety can be enhanced by an MBCP even in situations where a perpetrator does not change his behaviour—for example, if the program in the short term is able to help contain the risk a perpetrator poses to their partner and family, and over the longer term support a victim’s/survivor’s wellbeing.

The focus on the safety of women and their children as the core priority of all MBCPs also encompasses wider issues of risk and ethical considerations in evaluation, which will be addressed in the following sections.

Other considerations regarding measuring outcomes

Measuring outcomes for victims/survivors

Program evaluations of MBCPs should incorporate a dedicated evaluation stream focusing on the partner and family safety contact component of the program, including data obtained from victims/survivors about their experiences of partner and family safety contact and the impact it has (or hasn’t) had. It is important that outcomes measures focusing on victim/survivor safety do so in ways that are not solely linked to changes in the perpetrator’s behaviour. Victim/survivor safety can be enhanced by an MBCP even in situations where a perpetrator does not change his behaviour—if, for example, the program in the short term is able to help contain the risk a perpetrator poses to their partner and family, and over the longer term support a victim’s/survivor’s wellbeing.

MBCP evaluations can include a focus on questions about the ability of the program to identify and respond appropriately to new or escalated perpetrator-driven risk. In some instances where a program is unable to change a perpetrator’s attitudes or behaviours, managing the risk he poses to adult and child victims/survivors can be an important outcome. This includes the extent to which the program is able to collaborate effectively with other agencies towards a multi-agency risk management response.

Systems-level outcomes

Evaluating behaviour change outcomes can include the ways in which the program is intended to exert an influence on the integrated DFV response system of which

it is a part. Some program logics in the MBCP field therefore divide immediate and intermediate outcomes into two sections—one that represents systems-level outcomes, and the other at the program level.

Immediate systems-level outcomes are important to consider in the course of process evaluations. Examples include:

- agencies refer appropriately into the program
- referring agencies actively collaborate with the MBCP provider during the man's participation in the program
- agencies appropriately and proactively share and exchange information with the MBCP provider to help to identify, assess and manage perpetrator-driven risk.

Intermediate systems-level outcomes can include:

- the MBCP provider and partner agencies involved in an integrated response understand their own, and each other's, roles and responsibilities in engaging effectively with perpetrators and in responding to perpetrator-driven risk
- partner agencies and other stakeholders develop a nuanced understanding of MBCPs and adopt realistic expectations about their effectiveness
- the MBCP assists child protection and intensive family support services to hold perpetrators responsible for their impacts on children's welfare and family functioning and to ally with the non-offending parent
- the MBCP contributes to the ability of the integrated DFV response to manage high-risk, high-harm perpetrators.

Of course, there are many things that can influence intermediate (and ultimate) systems-level outcomes, beyond the influence of the MBCP. It is important to state these outcomes in terms of what the MBCP may impact.

Measuring ultimate outcomes

Many issues specific to MBCP evaluation complicate the likelihood of determining the achievement of "ultimate outcomes" (Vlais & Green, 2018), such as maintaining contact with victims/survivors over the long term when they are no longer partnered with the perpetrator, and maintaining contact with perpetrators once they have completed the program. This difficulty was illustrated by an evaluator we interviewed:

In most programs, the follow-up numbers of men that will respond to a questionnaire after a program is completed is below 10 percent ... We follow up with all men that agree to participate in the evaluation (usually just over 50%), but less than 10 percent of that cohort respond after 6 months.

As a result, many MBCP evaluations focus more on the *immediate* and *intermediate outcomes* of the program rather, than on *ultimate outcomes*. While this has long been viewed as problematic emerging research suggests validity in using robust measures to assess shorter-term change (Semiatin, Murphy, & Elliott, 2013; Silvergleid & Mankowski, 2006). These immediate and intermediate outcomes could include changes in perpetrators' attitudes and behaviour, assessments of perpetrator risk and accounts of victim/survivor autonomy and wellbeing. Some researchers argue that

measuring more immediate outcomes shows progress towards ultimate outcomes of a program logic (Silvergleid & Mankowski, 2006; Vlasis & Green, 2018). Thus, while it is important to continue to attempt to assess ultimate outcomes, immediate and intermediate outcomes can be used as indicators of ultimate outcomes.

Determining realistic outcomes of MBCPs

Recent increases in funding provision for MBCPs have meant that program providers are under greater pressure to produce adequate evidence of program effectiveness (Vlasis & Green, 2018). At the same time, there is growing acknowledgement in the sector that for most perpetrators, perpetrator engagement in a single MBCP is unlikely to lead to long-term and sustained behaviour change (Dutton & Corvo, 2006), and that MBCPs need to be viewed within a wider system of supports and interagency measures that aim to reduce violence against women and keep a perpetrator “in view”. Recent shifts in policy and practice highlight the importance of multi-agency information sharing during and following MBCP completion. Service knowledge of perpetrator behaviour, motivation to change, and potential risk informs decisions of other agencies and ensures that perpetrators are kept “in view” through integrated systems (No to Violence, 2006; Respect, 2017).

An important evaluation question to consider: Unintended outcomes

It is possible that unintended outcomes arise from an MBCP, which may be positive (additional value-added benefits) or negative (harmful). For example, one unintended outcome of psychoeducation on types of abusive and controlling behaviours may be that MBCP participants actually “learn” about and take on new problematic behaviours. Other types of unintended negative consequences of MBCP activity unfortunately appear reasonably common: some perpetrators attempt to use their participation in a program to manipulate service system and community responses for their own benefit (Opitz, 2014; Vlasis & Campbell, 2019), or distort what is covered by the program to extend control over their partner (Wistow, Kelly, & Westmarland, 2017).

Evaluation stakeholders, such as program facilitators, who are sometimes able to observe outcomes directly might be in a good position to pick up unintended outcomes like this. It is important, therefore, to ensure that these stakeholders are given the opportunity to discuss outcomes beyond those the evaluation is designed to measure.

Measuring unintended outcomes might be done in a direct way, such as by asking victims/survivors directly if there have been any additional negative or positive outcomes or consequences of the program after asking other evaluation questions. Another way of identifying positive and negative unintended outcomes is to use the “most significant change” technique (Davies & Dart, 2005) discussed later in the guide. Examples are shown in Table 4.

TABLE 3:

Examples of questions and indicators for assessing unintended consequences of MBCPs

Example evaluation questions	Example indicators
<ul style="list-style-type: none"> • What are the ways in which men's participation in the MBCP could inadvertently cause harm to family members? • Is identification of these risks and impacts a customary focus in family safety support work? • How are these risks and impacts minimised? 	<ul style="list-style-type: none"> • Documented occurrences of men using program participation as a tactic of control in the family safety support contact recording tool or risk management register/matrix • Documented occurrences of other harms connected with men's participation determined by auditing a representative sample of case files or completion reports to referrers • Practice guidance and prompts in the program manual for facilitators to monitor, identify and respond to harm connected with MBCP participation

Transferability

Transferability is "the extent to which the measured effectiveness of an intervention could be achieved in another setting" (Rychetnik, Frommer, Hawe, & Shiell, 2002, p. 119), including in a different place, organisation or target group. Whether evaluation findings are transferable depends largely on MBCP and context similarities. In our interviews with MBCP evaluators, the difficulty of generalising evaluation results was raised. Evaluators stressed how substantially different MBCPs are in terms of approach, program length, practitioner skill level, organisational support for the program, degree and depth of partner contact, and integration with the broader system, making MBCP evaluation findings unlikely to be transferable.

To assess whether a program will be effective in a different setting, it can be useful to ask the questions outlined in Box 3. Transferability is important to consider when deciding whether to implement a program that has been tested in another setting and when reporting evaluation findings for others to make decisions about transferability of your MBCP.

BOX 3:

Assessing transferability of MBCPs

To determine to what degree the results of one evaluation might be transferable to another, it is useful to ask the following questions:

- What is the level of DFV in the community in which you will implement the MBCP? How does it compare to the communities exposed to the primary MBCP?
- Are characteristics of men to be involved in the program similar to those in the program you want to transfer? For example, are they of similar cultural backgrounds and socioeconomic status? Is it possible that differences in characteristics will affect MBCP implementation or effectiveness? What adaptations might be made?
- Are the resources available to implement the MBCP similar to those for the program you want to transfer? For example, is there a similar organisational environment in terms of the partner agencies available to deliver an integrated response? Is there the same or a similar level of organisational support?

Adapted from Wang, Moss, and Hiller (2006, p. 79)



Replicability considers whether an MBCP—and which of its parts—should be implemented elsewhere. This may have important funding implications.

Additional information about transferability can be found here: <https://academic.oup.com/heapro/article/21/1/76/646412> (Wang et al., 2006).

A useful framework for thinking further about replicability can be found here: <https://cdn.catch-22.org.uk/wp-content/uploads/2015/11/Realising-Ambition-Programme-Insight-Evidence-is-Confidence.pdf> (Catch-22, n.d.).

Answering the evaluation questions

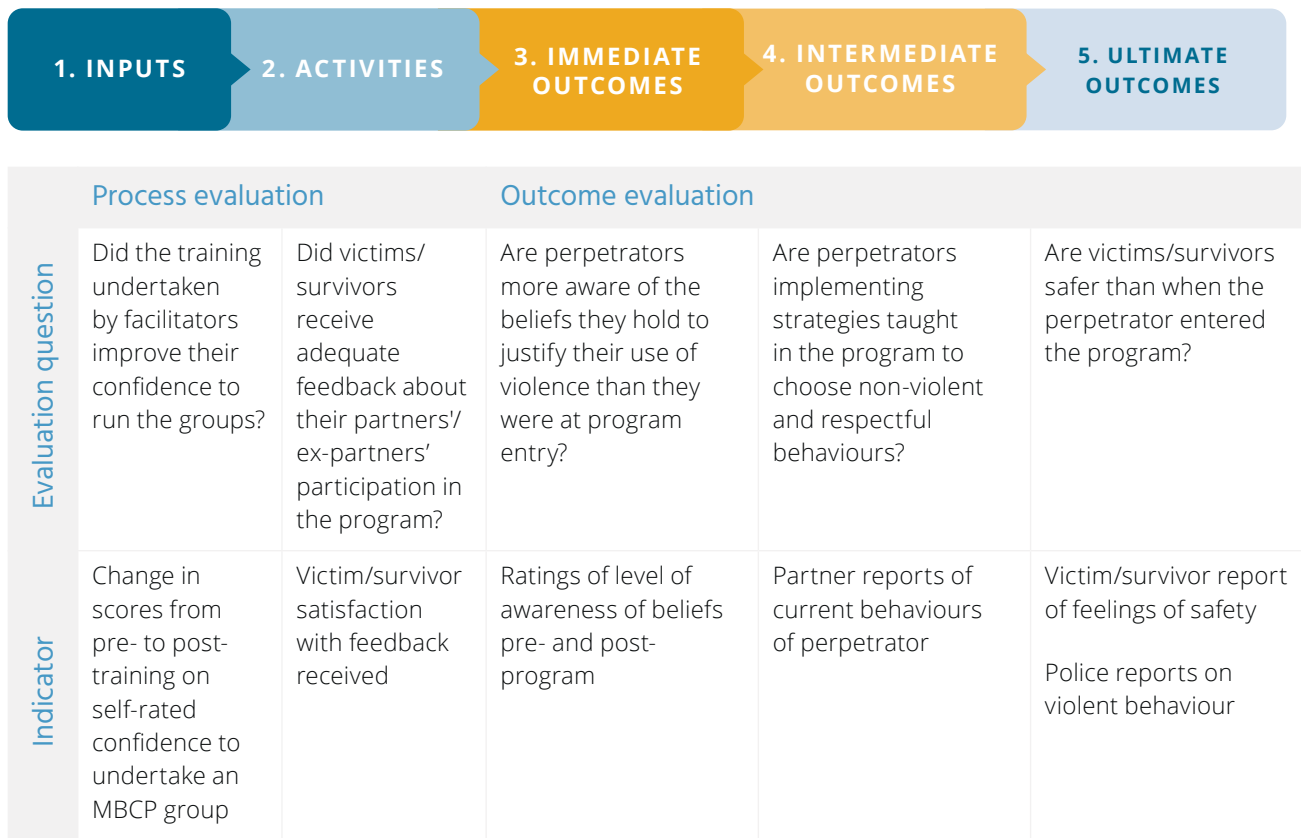
Choosing key indicators to answer your evaluation questions

Once a program logic and evaluation questions are decided, it is time to consider *how* these evaluation questions will be answered. To do this, you need to consider key indicators for each of the evaluation questions. An indicator is “a specific, observable, and measurable [marker of] accomplishment or change that shows the progress made toward achieving a specific output or outcome in your logic model” (Salabarría-Peña, Apt, & Walsh, 2007, p. 175). Generally speaking, indicators can be quantitative or qualitative. Most frequently, evaluations of complex programs, like MBCPs, use both types of indicators.

Quantitative indicators are reported numerically and can measure the scale of changes produced through the program (e.g. how much has changed and how many have changed). Qualitative indicators tend to describe “meaning and subjective experience” (Mark, Henry, & Julnes, 2000, p. 161), allowing you to answer questions unrelated to quantity or scale. For example, the evaluation questions “What are barriers to program delivery?” or “How can this aspect of the program be improved?” may result in the following indicators: “facilitator perceptions of how the program can be improved” or “facilitator views on barriers to perpetrators taking responsibility for their behaviour”. The indicators you select in your evaluation should contribute to answering your evaluation questions. Figure 6 shows some possible indicators that could be used to answer some of the process and outcome evaluation questions we derived from the program logic earlier in the guide.

FIGURE 6:

Indicators to answer evaluation questions arising from the program logic in Figure 3



Choosing quality indicators: SMART indicators

There will likely be many possible indicators that can be chosen to answer your evaluation questions. One way to assess the quality of proposed indicators is to assess whether it is a SMART indicator: specific, measurable, achievable, relevant and time-bound. Table 5 shows the example of using a facilitator's self-report of confidence to conduct the MBCP sessions as an indicator, which might be used before and after training.

TABLE 4:
How to assess indicators against the SMART criteria

SMART	Criteria	Meets criteria?
Specific	Does the indicator relate directly to the desired outcome?	✓
Measurable	Could we repeat the measurement? Do we have the resources required to undertake the measurement?	✓ Yes, though self-reported confidence is subjective and may vary over time
Achievable	Is the expected change achievable as a result of the intervention?	✓ If set reasonable expectation about how much confidence will increase
Relevant	Does the indicator reflect the expectations of stakeholders?	✓
Time-bound	Could the expected change happen in the measurement period?	✓ Could also include follow-up assessments to identify additional training needs after facilitation begins, as confidence may change once the groups commence



This information is adapted from “What makes a good performance indicator”, found here: <https://coastadapt.com.au/how-to-pages/identifying-indicators-monitoring-and-evaluation> (CoastAdapt, 2017).

Note that you can have more than one indicator for each evaluation question. In fact, it is often advisable to use more than one indicator to answer the same evaluation question. This is called data triangulation.

Data triangulation

Data triangulation in evaluation means combining different indicators to answer the same evaluation question. Triangulation can involve using different methods to get a better view of the answer to a particular question (also sometimes called “mixed methods”). For example, to measure “reduction in violent behaviour”, an evaluator might access police records and also ask victims/survivors about a perpetrator’s

frequency of violent behaviours. Data triangulation is also using the same data collection method with different points of view. For MBCPs this will likely involve getting the perspective of the perpetrator, victim/survivor, facilitators, and/or other case workers.

Problematic indicators requiring triangulation

Historically, there have been commonly used indicators in MBCP evaluations that are now considered less valuable, particularly when used alone. It may be worth considering why these are no longer considered sole indicators of MBCP outcomes and why they need to be triangulated with alternative indicators.

Program completion

Program completion or retention has often been used as an indicator of “success” in MBCPs, particularly by funders and commissioners of research. However, while retention remains a key concern in MBCPs generally, merely completing a program cannot account for change in perpetrator behaviour (Westmarland et al., 2010).

Perpetrator self-report

Frequently, comparisons are made between perpetrators’ self-reported data “pre-program” and “post-program” as an indicator of the program’s outcomes (Day & Casey, 2010). The use of self-report alone to measure change is problematic and unreliable for MBCPs. There are often vast differences in self-reports by perpetrators and ratings by victims/survivors, which raise concerns about the accuracy and usefulness of perpetrator self-reported data. As one of the evaluators we interviewed noted:

One of the really big challenges is that men will consistently self-report a huge amount of [positive] change, but their partners will not. So, what feels like a large change for them isn’t necessarily seen the same way by those affected by their violence. Given these issues, all measurements of change and safety need to be made through family [victim/survivor] contact.

The victim/survivor should always be given the opportunity to provide feedback regarding her own experiences of safety and wellbeing rather than relying on perpetrator self-reports of perceived change.

Recidivism data

Recidivism data is not considered a good indicator of victims/survivors’ and children’s safety as it fails to capture perpetrators’ use of a range of tactics to control women and their children. These tactics include emotional and financial controls and sabotaging women’s relationships with their support networks, all of which MBCPs ultimately aim to reduce (Vlais & Green, 2018; Walby et al., 2017). Recidivism is also not an accurate measure of attitudinal change (Kelly & Westmarland, 2015), as recidivism data might measure non-MBCP-related behaviour change or miss non-criminal forms of DFV. Yet many published evaluations of MBCPs have relied on recidivism data as the main outcome measures for program effectiveness, including data on re-offending (both DFV-specific and general criminological re-offending) and the time period to next re-offence. This reliance on recidivism data reflects the predominance

of corrections programs in the published MBCP literature (particularly in the United States), where victim-centred data has not been collected. While recidivism measures can in some circumstances have a role as part of a broader suite of measures, it is not recommended that they be sole measures of MBCP success.

Choosing data collection methods to answer the evaluation questions

Choosing data collection methods and overall study design is a key part of planning an evaluation, as it is essential to the collection of high-quality data, and to better enable the evaluator to infer a link between the activities of the MBCP and its outcomes. The choice of data collection method and evaluation design will be influenced by the evaluation questions and indicators selected, the intended use of evaluation findings, and users' confidence in the findings, as well as other contributing factors such as resources available for evaluation.

Some examples of quantitative and qualitative data collection methods and their advantages and disadvantages are shown in Table 6. Figure 7 provides examples of possible data collection methods for the indicators we outlined in Figure 6.



TABLE 5:

Advantages and disadvantages of various data collection methods

Method	Advantages	Disadvantages
Quantitative methods		
Questionnaire/ survey questions asking closed- ended questions (e.g. "how many times in the last week" ...); or ratings (e.g. rate from "very bad" to "very good"); or using psychometric tests	<ul style="list-style-type: none"> • Can be anonymous • Inexpensive and time-efficient • Can use validated measures that already exist • Can yield large sample sizes and substantial data to build evidence base 	<ul style="list-style-type: none"> • Provide information on "how much" but not "how" or "why" • Issues with partial completion • Can be an impersonal way to collect sensitive personal information • May need assistance of statistician or expert to perform the analysis • Some psychometric measures (i.e. validated questionnaires) need to be administered by a qualified person and there can be a cost for access • Need to ensure safety of victim/survivor respondents (e.g. might be risky to send via email or post, making distribution and collection more difficult)
Interviews with closed-ended questions (e.g. asking questions beginning "how many times ...")	<ul style="list-style-type: none"> • Face-to-face engagement with participants can facilitate rapport-building and result in more accurate data • Can provide explanation and clarification of questions to improve data accuracy and richness • May be more suitable than written methods for gathering sensitive information, as in an MBCP evaluation 	<ul style="list-style-type: none"> • Expensive • Time-intensive • Limited number of interviews you can conduct and therefore how much data can be collected
Observations using ratings, scores, checklists	<ul style="list-style-type: none"> • Can record program operation in real time and make necessary changes • Can reduce response biases of self-reports through observation by an independent third party • Direct observation can result in more accurate data and provide useful context 	<ul style="list-style-type: none"> • Checklists, ratings and scores may not capture important contextual information • Being observed can influence behaviours of program facilitators and participants • Expensive • If using multiple observers, will require training and guidance to yield reliable data (i.e. to achieve inter-rater reliability or consistency of ratings across observers)
Collation of administrative data (e.g. attendance records, recidivism data)	<ul style="list-style-type: none"> • Uses standard data and therefore does not place additional burden on program facilitators and administrators or evaluators • May have been collected over a long period of time, so could be used to test effects of new programs or changes to programs by looking at the data before and after 	<ul style="list-style-type: none"> • Gaining ethical approval to use administrative data can be difficult due to privacy concerns • Can be unclear how accurate such records are, as they are not collected by evaluators directly • Administrative data collected might not be useful for evaluation purposes, as it is recorded for other reasons

Method	Advantages	Disadvantages
Qualitative methods		
Questionnaires, surveys asking for comment, opinion or description (e.g. “please describe ...”)	<ul style="list-style-type: none"> • Can provide context to quantitative data or survey questions (e.g. “Please explain why you gave the rating above. What could we do to improve your rating?”) • Can be anonymous • Inexpensive • Can yield large sample which can then be analysed to identify common themes across responses 	<ul style="list-style-type: none"> • Can be impersonal, particularly if you are requesting sensitive information • Need to ensure safety of victim/survivor respondents (e.g. might be risky to send via email or post, making distribution and collection more difficult) • Wide variations in responses can be difficult to interpret • Can only provide minimal detail
Documentation review	<ul style="list-style-type: none"> • Comprehensive information (e.g. meeting minutes and facilitator manual) • Data can be retrieved from existing resources • Does not interrupt the program activities 	<ul style="list-style-type: none"> • Information may be incomplete or out-of-date • Restricted in terms of reviewing information that already exists (e.g. facilitator manuals might not be detailed enough to understand the activities of the program)
Individual interviews that ask open-ended questions	<ul style="list-style-type: none"> • Can gather in-depth information about “how” and “why” • Promotes engagement with the participant, which may yield better quality data than other methods and may be more suitable than written methods for gathering sensitive information • Can provide opportunity for victim/survivor to have her story heard, perhaps for the first time 	<ul style="list-style-type: none"> • Time-consuming to schedule, conduct and analyse • Difficult to analyse and draw causal conclusions from the data • Costly • May be prone to interviewer bias • Victim/survivor may be unable to disclose current risks, etc.
Case studies	<ul style="list-style-type: none"> • Can collect and comprehensively examine all aspects of participant experiences • Engaging data source through which to illustrate participant experiences to evaluation users • Allow for cross-comparison of cases 	<ul style="list-style-type: none"> • Time-consuming to collect and produce • Case studies might not be generalisable or transferable, especially since MBCPs widely differ
Observations that result in descriptions/ qualitative feedback	<ul style="list-style-type: none"> • Can record program operation in real time and make necessary changes • Can reduce response biases of self-reports through observation by an independent third party • Direct observation can result in more accurate data and provide useful context 	<ul style="list-style-type: none"> • Can be difficult to interpret and categorise observations • Being observed can influence behaviour of program facilitators and participants • Expensive
Focus groups	<ul style="list-style-type: none"> • Enable common issues and themes to be discussed • Efficient in terms of time and expense compared with individual interviews • Can provide a range of responses on key issues 	<ul style="list-style-type: none"> • Difficult and time-consuming to analyse responses • Requires a trained facilitator • Difficult to schedule due to large number of participants • Not suitable for the discussion of sensitive information • Confidentiality issues—may not be appropriate to ask about details of experiences of DFV in a group setting

FIGURE 7:

Possible data collection methods used for indicators outlined in Figure 6



	Process evaluation		Outcome evaluation		
Evaluation question	Did the training undertaken by facilitators improve their confidence to run the groups?	Did victims/survivors receive adequate feedback regarding their partners'/ ex-partners' participation in the program?	Are perpetrators more aware of the beliefs they hold to justify their use of violence than they were at program entry?	Are perpetrators implementing strategies taught in the program to choose non-violent and respectful behaviours?	Are victims/survivors safer than when the perpetrator entered the program?
Indicator	Change in scores from pre- to post-training on self-rated confidence to undertake an MBCP group	Victim/survivor satisfaction with feedback received	Ratings of level of awareness of beliefs pre- and post-program	Partner reports of current behaviours of perpetrator	Victim/survivor report of feelings of safety Police reports on violent behaviour
Possible data collection methods	Questionnaire completed before and after the training that includes a rating scale of confidence to run an MBCP group (quantitative)	Interview with victim/survivor in which she is asked to rate satisfaction from "not at all satisfied" to "completely satisfied" (quantitative), and to answer a question asking what additional feedback she would have found useful (qualitative)	Intake worker rating of perpetrator awareness of beliefs (quantitative) Facilitator rating of awareness of beliefs post-MBCP (quantitative)	Victim/survivor interview (qualitative)	Interviewer-administered questionnaire of victim-/survivor-rated perceptions of safety (quantitative) Document review of police reports (qualitative)

Validated scales

One of the data collection methods listed above is the use of psychometric tests or validated scales. A difficulty in evaluation is identifying appropriate measures to assess the impacts of an MBCP (Day et al., 2019). In our consultation with MBCP evaluators, there was consensus that both understanding of, and access to, appropriate outcome measures represented a significant roadblock to evaluation. Other issues such as time and funding constraints meant that validated outcomes measures to assess the impact of an MBCP were rarely used.

In Appendix B of the full research report, we have included a table that outlines a number of validated measures across the three key outcome domains of:

1. long-term changes in perpetrators' violent and controlling behaviour
2. adult victims/survivors' safety, wellbeing and freedom
3. children's safety, wellbeing and family functioning. (Adapted from Respect, 2017)

While there are few measures suitable to use alone for the purposes of MBCP evaluation, the table in Appendix B provides guidelines on which instruments, or combinations of instruments, may be valuable for use in an MBCP evaluation. Information is also provided on access, costs and/or administration requirements associated with each measure.

Most Significant Change

One lesser-known method of qualitative data collection worth noting is the Most Significant Change (MSC) technique. MSC can be especially useful for illustrating to different groups of stakeholders what "success" looks like from other perspectives, for uncovering organisational values and for identifying "unintended outcomes" that were not anticipated in the program logic (Davies & Dart, 2005). MSC is a useful addition to other techniques of measuring intermediate outcomes and feeding them back to stakeholders throughout the program.

Using the MSC technique involves asking those "in the field", such as facilitators of MBCPs, a question like "What was the most significant change that took place for the men in your program over the last month?" as well as why they think that change was "most significant". This question might be asked at regular intervals. The responses are then allocated to pre-determined "domains" or themes that the program was designed to assess (e.g. changes in understanding, parenting, or communication). Davies and Dart (2005) refer to a process of collating and summarising these responses through consecutive stages with different groups of people within an organisational line management structure.



This process requires substantial time and investment but may be suitable to consider for your evaluation. A guide to the MSC technique has been written by the creators of the technique and can be found here: <https://www.mande.co.uk/wp-content/uploads/2005/MSCGuide.pdf> (Davies & Dart, 2005).

An overarching consideration: Reducing the burden of data collection

A major challenge in MBCP evaluations is the additional burden placed on program participants, facilitators and others to provide additional data beyond that which the program provider routinely collects. Considerable data and risk-related information is already collected from program participants (perpetrators and victims/survivors) in the course of initial and ongoing risk assessment and case reviews. If not sensitively and carefully planned, introducing new data collection requirements on top of existing assessment processes can be onerous both for program participants and practitioners. One approach for an internally conducted evaluation is to embed data collection in a program's initial and ongoing assessment processes.

One example of how this can be done comes from Project Impact, arising from the Work with Perpetrators European Network (WWP-EN, 2019). Project Impact developed tools that served the dual purposes of initial assessment templates and program evaluation outcome measures. WWP-EN developed separate tools to use with perpetrators and their (ex-)partners that are administered at different time points (pre-program, mid-program, post-program and follow-up) and capture use/experience of coercive controlling violence, victim/survivor safety, hopes for the relationship, and wellbeing of children. Provided that the particular items and question wording were not altered, these tools could be merged into the program's customary initial assessment, risk assessment and program review templates. They could then be administered by program practitioners, so that the same items contribute towards both program clinical and program evaluation goals.

Selecting an appropriate and achievable study design

Types of study designs

A plan for measuring outcomes will need a description of the study design. There are many possible study designs, though relatively few have been used in formal MBCP evaluations due to the real-life constraints of conducting evaluation in this context. The type of design you choose will depend to some extent on how much you want to be able to determine "causality".

The extent to which you can attribute your findings to the MBCP and not to other external factors may be referred to as "causality". In outcome evaluation, it is important to discern whether your program activities led to the measured outcomes, rather than these outcomes being the result of other external factors. This is especially true as MBCPs take place in the context of other perpetrator interventions and integrated response systems. However, in process evaluation, you might just want to understand stakeholders' experiences of the program, rather than being concerned about changes produced. In this case, causality is not a primary concern.

Designs that are best at determining causality often involve randomly assigning participants to either an intervention or control group (e.g. randomised controlled trials). Randomly assigning participants to either group means that both groups are assumed to be "equivalent" (e.g. in terms of sociodemographic variables and relevant history). One group receives the MBCP (i.e. intervention group) and one does not

(i.e. control group). This allows the evaluator to conclude that any differences in outcomes between the two groups are due to the intervention (Tharp et al., 2011). However, for ethical and practical reasons, this design is rarely used in MBCP evaluation (see next section: Experimental rigour vs. ethical considerations).

Evaluation designs can vary in many ways, including whether participants are randomised to a control group, whether there is a control group, and at how many time points outcomes are measured. These design components can also be combined in different ways to create numerous variations of study design, each with a varying ability to determine causality. For example, a non-equivalent control group design is one involving a “natural” control group where participants are not randomly assigned to intervention or control, but outcomes are compared (see Box 3 for an example). Pre-test to post-test design compares outcomes before and after the intervention to determine if change has occurred. Sometimes a control group is used and compares pre–post changes between groups. Longitudinal designs measure outcomes before and after the intervention but also at one or more time points following completion (see Box 4.) One sub-type of longitudinal design is time series design, where outcomes are measured at multiple time points before and after the intervention.

There are other types of study design that do not focus on causality, for example cross-sectional designs in which a measurement is taken at a single point in time (e.g. a measure of attitudes toward women completed by perpetrators in the MBCP to determine the most common attitudes in MBCP participants). Cross-sectional designs might also assess the correlation between two concepts (e.g. scores on an attitudes toward women scale and [ex-]partners’ ratings of violence) at a single point in time. This type of correlational design is limited in its ability to determine that one construct “caused” the other because it doesn’t rule out other influences and cannot determine the direction of the relationship (i.e. is violence the result of poor attitudes or do poor attitudes develop as a way to justify violence?). However, these designs may also be useful if they assist in answering your evaluation questions, depending on what they are.

Examples of various study designs, and their relative advantages and disadvantages, are described in more detail in Table 7.



TABLE 6:

Examples of types of study design

Study design	Description/example	Strengths	Weaknesses
Randomised controlled trials (RCTs)	RCTs involve the random allocation of participants into intervention and control groups (i.e. some participants will partake in the MBCP and others will be allocated into a no-intervention “control” group). The two groups are compared on outcomes	<ul style="list-style-type: none"> • Considered the “gold standard” test of effectiveness • More likely to produce evidence that outcomes are due to the MBCP rather than other factors, such as maturation, demand characteristics, or other possible mechanisms of change which participants might be exposed to (Tharp et al., 2011) 	<ul style="list-style-type: none"> • Most resource-intensive • May not be ethically appropriate for use in DFV contexts, particularly given that random assignment may put women victims/survivors at risk of further harm and psychological distress (Arai et al., 2019) • May be impacted by a range of intersecting support systems (e.g. support for victims/survivors from other integrated response agencies) which may reduce the likelihood of “controlled” experimental contexts for MBCP evaluation (Howarth et al., 2019, p. 60)
Non-equivalent control groups design	Trials using a control group and intervention group where participants are not randomly assigned to either group (e.g. “control” participants in a different geographic location where MBCP will commence soon, but outcomes are measured before they attend the MBCP)	<ul style="list-style-type: none"> • An alternative when you are unable to randomly allocate participants • Enables comparison across groups and/or across time points 	<ul style="list-style-type: none"> • Differences between comparison groups may confound the results (e.g. could their different geographic location make them different to the intervention group in a way that will affect the outcomes?). See Box 3 for Project Mirabal (Kelly & Westmarland, 2015) • May be ethically inappropriate given that the “control group” will be men who use violence, or men who are on a “waiting list” for the MBCP and should participate in an MBCP as soon as possible • There are further ethical issues in allocating a “waiting list” of men for MBCPs as there may be immediate and long-term risks to women
Single group pre-test–post-test design	Also called “single-group” or “within subjects” design; involves collecting data in one group, who are compared before and after the MBCP	<ul style="list-style-type: none"> • Simple design • May be used when comparison groups are not available or ethically inappropriate 	<ul style="list-style-type: none"> • Limited ability to infer causality (i.e. that changes were the result of the MBCP)

Study design	Description/ example	Strengths	Weaknesses
Longitudinal design (see Box 4 for a broad example of longitudinal design)	Measures of outcomes are repeated over a long period of time (e.g. pre-MBCP, post-MBCP, 12 months follow-up and 24 months follow-up)	<ul style="list-style-type: none"> • Allows for measurement of long-term maintenance of change in attitudes and behaviours (i.e. sustained outcomes) following the MBCP • Could identify when positive outcomes are lost and therefore when repeat intervention might be needed 	<ul style="list-style-type: none"> • Access to perpetrators and (ex-) partners is very difficult following completion of the MBCP. This requires an ability to get in contact and the willingness of evaluation participants to stay involved • Time- and resource-intensive to track participants over time (see Box 5 for Brown, Flynn, Fernandez Arias, and Clavijo, 2016)
Case control	Case control (post-intervention only) retrospectively compares data between intervention and non-intervention groups (e.g. compare recidivism data for perpetrators who did and did not participate in an MBCP)	<ul style="list-style-type: none"> • May be used when baseline data is not available and for descriptive study • Might be able to use a large retrospective dataset to make comparisons between MBCP and non-MBCP groups 	<ul style="list-style-type: none"> • Limits to determining causality • Difficult to determine retrospectively what external factors might have caused differences in outcomes between those allocated and not allocated to MBCPs (e.g. were some perpetrators historically not considered “suitable” for MBCPs and are the reasons and the data supporting this still available?) • Consistent data would need to be available to allow collation of the dataset over time
Cross-sectional design	Measures a construct at a point in time in order to describe it (e.g. prevalence of particular types of controlling behaviour among men who enter MBCPs), or measures two or more constructs and correlates them (e.g. does holding particular attitudes relate to particular types of controlling behaviours?)	<ul style="list-style-type: none"> • Can be quick and easy as measures a single point in time • Can be useful describing how common a construct is (e.g. a particular attitude or type of controlling behaviour, victims'/ survivors' level of wellbeing at the beginning of perpetrators' entry into the MBCP) • Can be useful to generate further hypotheses for testing • Can collect information on multiple constructs at once (e.g. attitudes and behaviours) 	<ul style="list-style-type: none"> • Cannot determine causality nor directions of relationships between constructs

BOX 4:**Non-equivalent control group design: Domestic violence perpetrator programmes—Steps towards change (Kelly & Westmarland, 2015)**

Project Mirabal had a unique approach to providing a “non-intervention” comparison group (non-equivalent/non-randomly assigned control group) for comparison of outcomes between those attending and not attending a perpetrator program. The project attempted to compare outcomes from female (ex-) partners of men attending the perpetrator intervention with those whose partners were not attending because no program existed in the area. The women in the intervention and comparison groups were matched on “basic demographics, length of relationship and baseline levels of violence and abuse” (Kelly & Westmarland, 2015, p. 7).

The project illustrates one of the major challenges of a non-equivalent comparison group design: the two groups differed substantially in important ways. The comparison group were more likely to have children not in contact with their father, and they were far more likely to still be partnered with the perpetrator than the intervention group. These alone could explain differences in the presentation of outcomes and might reflect that the women were “at different points in the process of dealing with domestic violence” (Kelly & Westmarland, 2015, p. 8). Differences in life circumstances might also affect men’s motivations for change. Consequently, Kelly and Westmarland (2015, p. 8; see also Kelly et al., 2013) concluded that

whilst we do have comparison group data [...] the fact that they are not an equivalent comparison group rendered the comparative data difficult to interpret in a way where we could be sure of our explanations [...] If we had much higher numbers of men going through [domestic violence perpetrator programs] and higher numbers of research participants, it would have been possible to control for these differences. However studies of this nature do not tend to recruit the numbers that would have been required, and developing appropriate comparison or control groups unfortunately remains methodologically problematic.



BOX 5:

Longitudinal evaluation: European model for MBCP evaluations

Lilley-Walker, Hester, and Turner (2018), in a review of the methodologies and measures used in European (and United Kingdom) evaluations of MBCPs, proposed a multi-point, longitudinal evaluation model with common elements of quality evaluations in the field. Based on their review of European MBCP evaluations, the authors concluded that studies need to

specify who exactly is participating, completing, and dropping out, at what point, and their motivations for doing so. Thus, careful attention must be paid to the types of information being collected—and also then reported—at different time points in order to better understand what and how behaviour and attitudes might change throughout the course of the programme. (Lilley-Walker et al., 2018, p. 880)

They recommend that the structure of a quality MBCP evaluation be summarised as follows:

T0 (pre-program)

- Size and type of sample at intake
- Referral routes and program pathways
- Excluded referrals and referral drop-outs

T3 (program end)

- Size and type of sample at end of program
- Measures
- Program completers

T1 (start of program)

- Size and type of sample at start of program
- Initial measures
- Excluded participants and drop-outs

T4 (follow-up)

- Measures
- Completers vs. non-completers

T2 (during program)

- Process/role and quality of facilitation and other program components
- Measures
- Drop-outs



BOX 6:**Challenges of longitudinal evaluation:****A study of the impact on men and their partners in the short term and long term of attending MBCPs (Brown et al., 2016)**

This well-known evaluation of MBCPs from 12 sites across Australia used a longitudinal design. This involved attempting to follow up with men participating in the program and their partners for 3 years, commencing at the start of the program to 2 years after the end, using surveys and interviews.

The evaluation report for this study illustrates one of the difficulties of longitudinal design: keeping in contact with participants over a long period and maintaining their participation in the evaluation. In the first wave of data collection on entry to the MBCP, there were 270 men in the study. The second wave at the end of the program attracted 110 men, but only after significant efforts to contact many of them. The third wave, 1 year later, included just 45 men, and their inclusion also required significant effort:

This number was gained after mail outs and emails, followed by phone calls to men [...] offers to pay for the completed surveys, further phone calls and further mail outs and emails. (Brown et al., 2016, p. 33)

Attempts to contact partners met with so little success that no quantitative data was able to be obtained.

Attempting to conduct any longitudinal evaluation should consider the significant effort and resources needed to ensure ongoing participation over a long period of time. Consideration should be given to what strategies can be used and therefore the resources required to encourage ongoing engagement in the evaluation. Furthermore, it is imperative that quality measures of more immediate outcomes are used, as discussed in the section “Measuring ultimate outcomes”.

Experimental rigour vs. ethical considerations

Most MBCP evaluations do not employ rigorous experimental design in the form of a randomised controlled trial (RCT).¹ While RCTs are considered the “gold standard” in evaluation, there is debate regarding the ethics and appropriateness of RCTs in MBCP contexts (Bender, 2017; Logan, Walker, Shannon, & Cole, 2008). Random assignment may put victims/survivors at risk of further harm and psychological distress because those in the control group do not immediately receive the intervention (Dutton et al., 2003). In addition, MBCPs take place in a broader system that influences perpetrator outcomes, significantly reducing the likelihood of “controlled” experimental contexts for MBCPs (Howarth et al., 2019, p. 60) as the MBCP is never delivered in isolation from other systemic interventions.

External factors affecting measured outcomes

Where it is not possible to use an evaluation design that can determine causality, and when causality is important, it can be useful to record factors external to the MBCP that might affect the measured outcomes. Such factors may include:

- Overall group differences: for example, one participant group may include a high number of high-risk, high-harm perpetrators, skewing the results because positive outcomes might be more difficult to achieve.

¹ RCTs are regarded as the most rigorous evaluation method because the random allocation of participants into different “treatment” groups (where some participants will partake in an intervention and others will be allocated into a no-intervention “control” group) provides evidence that the outcomes are due to the program itself, and not other possible factors.



- Individual differences: for example, participants of an MBCP group may be led by demand characteristics, such as a motivation to illustrate they have reduced their violent and controlling behaviour in order to see their children, but the change may not be maintained in the long-term.
- Outside factors: MBCPs likely represent one aspect in a range of multi-agency initiatives that aim to reduce violence against women. Men attending the program (and/or victims/survivors) may be receiving additional interventions such as supervision from a corrections officer. Similarly, victims/survivors may receive multiple services (e.g. safe housing, access to financial resources and legal advice) at the same time that their partner is completing the MBCP. These external services may influence the overall change recorded in pre–post MBCP measures.

Measurement of these “external” factors that might affect program outcomes can assist in better understanding the evaluation findings. For example, if the evaluation shows that participants with particular characteristics (e.g. men under 30 years of age) were most likely to drop out before program completion, then it might indicate, firstly, that a different approach is needed to maintain engagement of younger men, but also that the program outcomes might not be generalisable to younger men, since their intermediate outcomes (measured at program completion) were not measured.

Collective impact

Throughout most of the guide, we have referred specifically to measuring outcomes related to a single MBCP. However, it is worth noting a growing movement toward measuring “collective impact” (Kania & Kramer, 2011). Collective impact refers to “the commitment of a group of important actors from different sectors to a common agenda for solving a specific social problem” (Kania & Kramer, 2011, p. 36). Initiatives designed to have a collective impact require several key elements, not least of which are for participating organisations to have a common social agenda, a shared measurement framework and a shared plan of action (Kania & Kramer, 2011).

Collective impact initiatives for reducing the impacts of DFV within communities are increasing in Australia.² These initiatives require evaluation, and these evaluations are necessarily long-term and complex, as outcomes must be measured across the broad community they aim to target.

While the evidence base for collective impact as a strategy and the most accepted means of evaluating collective impact are still emerging, there are resources available if collective impact evaluation is something you want to know more about. A good place to start may be the Australian Institute of Family Studies’ webpage on evidence and evaluation for collective impact.³

2 See e.g. Cairns Collective Impact on Domestic and Family Violence <https://www.dvcairns.org/collective-impact-on-dv/>; Cardinia Shire Together We Can <http://togetherwecan.org.au/>; Logan Together <https://logantogogether.org.au/resources/>

3 See <https://aifs.gov.au/cfca/publications/collective-impact-evidence-and-implications-practice/evidence-and-evaluation>

Deciding who should conduct your evaluation

Evaluations vary widely in size and scope, and whether the evaluation is conducted by staff within the program or by an external evaluator will depend on a variety of factors like time, budget constraints, evaluation questions and study design, as well as ethical considerations. There are advantages and disadvantages to conducting both internal and external evaluations, and some of these are described in Table 9.

There is also the option to combine the internal and external evaluation components such that external evaluators are sub-contracted to conduct part of the evaluation. In one of the consultations for this guide, a program coordinator described employing an evaluator to observe and provide feedback on the facilitation of some groups. This is an example of when having an external evaluator would be most useful for providing an independent and objective view. Similarly, an external evaluator may be useful when particular expertise is required—for example, when a strong understanding of the cultural sensitivities of working with Aboriginal and Torres Strait Islander peoples is needed.

More information on internal and external evaluators can be found in the document, *A fundamental choice: Internal or external evaluation?* (Conley-Tyler, 2005). You can also complete the checklist to help make a decision about whether to do your evaluation internally or externally.⁴

An overarching consideration: Partnerships between program providers and evaluators

In some situations, working relationships between a program provider and an evaluator might be temporary or one-off. This can particularly be the case when the funder commissions an evaluation project for a new or existing program; the program provider might have no pre-existing relationship with the evaluator, and the working relationship ends once the evaluation is completed.

In other situations, opportunities can arise for program providers and evaluators to develop a longer-term partnership. While relevant applied research centres cannot fund program evaluations, industry and community partnerships and opportunities to conduct research that makes a difference can be highly valued.

These partnerships can help to strengthen a program provider's confidence about participating in evaluation and research activities. Although large agencies might have their own (still generally small) research unit and internal research ethics committee, in general, MBCP providers lack the knowledge and skills required to drive evaluation and research activity. Correspondingly, most applied research centres in the social or human sciences lack the subject matter expertise and understanding of some of the complexities and contentious issues involved in evaluating DFV services and programs. Partnership development—including researchers sitting in on MBCP group sessions and being involved in program planning and review activities—can help lay the foundation for quality evaluation activity once funding is in place to commence an evaluation process.

⁴ See https://www.aes.asn.au/images/stories/files/Publications/Vol4No1-2/fundamental_choice.pdf

TABLE 7:

Advantages and disadvantages of internal and external evaluators

Internal	External
+ive: Free or low-cost	-ive: Expensive
+ive: Evaluation can be implemented more quickly as facilitators and program managers are familiar with program and evaluation needs	-ive: Can require significant lead time before it begins. The evaluators need to be available, and then to spend some time understanding the program
+ive: Immediately useful as evaluators and users may be the same or in close proximity	-ive: The evaluator is not around in the longer term to ensure the results are utilised
+ive: Can capitalise on existing relationships to leverage participation of stakeholders in the evaluation (e.g. perpetrators, case workers)	-ive: It can be difficult for men who use violence and victims/survivors of DFV to trust in someone they don't know and for them to gain the willingness to talk to them for the purpose of evaluation
+ive: Facilitators as evaluators are already trained in avoiding collusion with perpetrators and are aware of the sensitivities of interviewing victims/survivors	-ive: Requires expertise in interviewing victims/survivors in ways that do not re-traumatise, and program participants in ways that avoid collusion with their violence-supporting narratives (see the section "Ethical considerations" for a more detailed discussion)
-ive: Participants in the evaluation (e.g. perpetrators) may be less likely to report negative thoughts or opinions of the program to internal evaluators in case they jeopardise their relationship with program staff	+ive: The evaluator's lack of involvement with the MBCP means stakeholders might be more likely to report things they don't like or that they don't think are working about the program. This can make the evaluation findings more useful
-ive: Participants can feel pressured or coerced into participating in an internal evaluation because of their existing relationships with those involved	+ive: Participants can feel less pressured or coerced into participating in the evaluation because they have no pre-existing relationship with the evaluator
-ive: Program staff might not have strong evaluation expertise	+ive: Evaluation consultants and organisations typically have a great deal of expertise and experience in a range of evaluation designs and methods. This means they might design a more comprehensive and useful evaluation, and that they might be able to carry out more complex evaluations than can be done by internal evaluators. These evaluation experts can also help to build the capacity of internal staff in the evaluation of their programs
-ive: Internal evaluators need to find extra time to do the evaluation. This might impact on the quality of the evaluation particularly if other demands are placed upon them	+ive: External evaluators are contracted to conduct the evaluation, and there would be less demand on the time of program staff to conduct evaluation activities
-ive: Users of the evaluation findings (e.g. program funders) may perceive internal evaluations as less trustworthy than external evaluations because they are seen as more likely to report favourable results (especially if there are funding decisions involved)	+ive: External evaluators are likely to be less invested in "positive findings" from the evaluation and are therefore likely to be more objective. They are also perceived as more objective by users, who therefore might see them to be more "credible" and thus have more faith in the findings

Note: "+ive" refers to positive and "-ive" refers to negative

Ethical considerations

Principles of ethical research

In Australia, all research (including evaluation activities) conducted with human participants must abide by the National Health and Medical Research Council's (NHMRC) *National Statement on Ethical Conduct in Human Research* (the Statement) (NHMRC, 2018). If you are planning to conduct an evaluation yourself, it is best that you read the Statement carefully, as it details how to adhere to a number of ethical research principles focused on participant protection.⁵

Broadly, the principles of ethical research involving humans are *research merit and integrity, justice, beneficence, and respect*.

In practice this means that, regardless of whether you are conducting your own evaluation or commissioning external evaluators, you must ensure that the evaluation design and processes address the following:

- how participants will be fairly and safely recruited and provided with enough information to give informed consent to participate. This will be especially true when involving women and children who are victims/survivors of DFV. Their participation must not place them at any additional risk
- how informed consent will be obtained
- what data will be generated or collected and how; how this data will be used and analysed, stored, disposed of and shared; risks associated with data collection; how the data collection adheres to the general ethical principles
- how the findings will be communicated to participants and to whom else they will be communicated and how, adhering to all the ethical principles
- how the findings will be communicated to a wider audience to ensure they contribute to broader knowledge in practice or the broader good
- how long the data will be retained after the project, how it will be stored to maintain confidentiality, and whether it will be available to others for future use.



Ethical guidelines specifically for evaluation

As well as these ethical guidelines for all research activities, the Australasian Evaluation Society (AES) has *Guidelines for the Ethical Conduct of Evaluations* (Australasian Evaluation Society, 2013). These guidelines encompass commissioning and preparing for an evaluation, conducting the evaluation, and reporting the results of the evaluation.⁶

Particular ethical issues to consider in evaluations of MBCPs

This section outlines some of the practical and ethical challenges evident in MBCP evaluation that may require special consideration.

⁵ See https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018#toc_95

⁶ See https://www.aes.asn.au/images/stories/files/membership/AES_Guidelines_web_v2.pdf

Involving perpetrators in evaluation

An overarching ethical principle of engaging both victims/survivors and perpetrators in MBCP evaluation is that they are positioned as active agents and can make informed decisions about their participation (Downes, Kelly, & Westmarland, 2014).

It is important to acknowledge that perpetrator engagement with an MBCP may intersect, and be motivated by, outside pressures and expectations. This is particularly relevant if a perpetrator engages in an MBCP while child contact and custody disputes are ongoing or if they are affected by civil/criminal proceedings. Their participation in the MBCP may also be linked to their motivation to maintain or resume a relationship with an (ex-)partner (Downes et al., 2014).

Perpetrators may also be reluctant to be involved in research and/or evaluation and have reservations about speaking openly about their current circumstances and past actions. Those engaging with perpetrators need to recognise potential risks but also acknowledge their involvement as voluntary participants and as men and fathers capable of positive change (Downes et al., 2014).

There are also particular circumstances where engagement with perpetrators has potential to harm. For example, perpetrators may use the evaluation process to rationalise and justify their use of violence against their (ex-)partner (Hearn, Andersson, & Cowburn, 2007). The possibility of collusion with the narratives and thinking that perpetrators use to minimise, deny and rationalise their use of violence should be acknowledged, and interviewers need to be adequately skilled or supervised to counteract such instances in the evaluation process.

There are numerous other potential risks in the interview process that need to be observed and mitigated in MBCP evaluation, including:

- risk of inadvertently reinforcing the perpetrator's position as a victim, or what is known in the MBCP field as his "victim stance" (e.g. "I'm glad this evaluator understands my situation and how unfair she has been to me, because the group facilitators sure don't!")
- risks associated with the provision of a private one-to-one interview setting. Perpetrator engagement and disclosure in the interview process may mean they avoid disclosing information in the more transparent, accountability-based, group setting
- the possibility of a perpetrator disclosing information in the course of the evaluation suggesting potential risk of ongoing or future violence. This includes the associated risk he poses to victims/survivors that the program practitioners might not be aware of
- risks associated with the perpetrator's reflection of the evaluation process. For example, the interview may incite problematic thinking that the facilitators aren't aware of (e.g. "This interview has reinforced for me how hard I'm trying to change my behaviour, and how ungrateful she is for my efforts").

The significant majority of perpetrators involved in an evaluation or research process will not pose a physical safety risk to interviewers. However, it is important for program providers to alert the evaluator when any such risk might arise. Risk mitigation strategies can then be put into place (e.g. a program practitioner being

present at the interview). It is also good practice for evaluators to follow general precautions, such as never interviewing a perpetrator in a building alone and having ready access to a duress alarm.

Evaluators and the program provider need to have protocols in place so that if the evaluator believes that the interview with the perpetrator has either in itself escalated risk, or that new information relating to risk was revealed, they will be directed to confer with the program provider to determine appropriate action, which could include contacting the victim/survivor about safety planning.

It is important for program staff of the MBCP being evaluated to consider what these risks could be and how to mitigate them. One way to mitigate such risk is through appropriate training and supervision of evaluators.

The importance of training and expertise in MBCPs

It is well recognised that MBCP delivery requires specialist skills and experience, including a strong understanding of the intersections of gender, behavioural factors, safety and cultural aspects which impact and support violence against women and children. It is imperative that evaluators have a good knowledge of the complexities in MBCPs and are well equipped to identify any issues as the evaluation progresses. This is particularly important if evaluators are engaging in qualitative interviews and individual consultations with men who use violence.

One of the evaluators interviewed for the study noted that because data from victims/survivors are often very important, it's crucial that the evaluators conducting these interviews have sensitive and sufficiently advanced interviewing skills and understand DFV. When an evaluator conducts an interview with a victim/survivor, it may be the first lengthy face-to-face discussion that the victim/survivor has had with a "professional". She might have a lot to tell, and it's critical that the evaluator does not try to limit her voice or the narration of her story if the victim/survivor is finding that the telling of her story is helpful. Similarly, interviewers need to be able to conduct interviews in ways that don't re-traumatise victims/survivors.

When you commence the process of ethical approval, you will also likely be asked to consider whether members of the evaluation team have the necessary knowledge and skills to conduct the evaluation.

A number of peak bodies operating at the state and territory level implement specialist training in DFV. For example, there are training and short courses that offer a basic understanding of the dynamics of working with perpetrators of violence (e.g. they address concepts such as responding appropriately to perpetrator invitations to collude; identification of risk escalation; perpetrator accountability; responsibility and victim/survivor safety; basic micro-skills in perpetrator engagement), and more complex specialist training for experienced practitioners.

The importance of training may also depend on whether the MBCP evaluation is being conducted by internal staff or external evaluators, and the level of expertise in the evaluation team. Internal evaluations are likely to engage staff or expert practitioners who have knowledge of the MBCP field. Internal staff are also expected to have DFV-

specific training, supportive organisational policies and procedures, and available supervision from experienced DFV practitioners. However, given the complexities involved in data collection in MBCPs, external evaluators require specific training when engaging with perpetrators for evaluation purposes. For example, workshops addressing incidents of possible collusion will assist external evaluators to effectively identify and respond to male perpetrators by learning how to identify and resist such invitations.

Management and collection of data from victims/survivors

There are specific considerations when collecting and managing data from victims/survivors related to recruitment, formal written consent, and data collection that require significant planning and an ongoing awareness of risks. Victim/survivor safety needs to be prioritised in the process. At the same time, the range of data collection methods should also attend to victim/survivor agency and advocacy. This includes ethical practice in data collection to ensure the evaluation does not encourage any of the following issues.

Placing victims/survivors at risk of further violence from their (ex-)partners

Victims/survivors need to be informed about aspects of the evaluation in order to voluntarily participate. However, evaluators also need to ensure that their methods of communication are not putting victims/survivors at undue risk of further violence and abuse. For example, if a victim/survivor remains in the household or in contact with her abusive partner, it may not be safe to send information about the evaluation to her home address (Downes et al., 2014). Equally, if a victim/survivor is requested to participate in a telephone interview, there may be particular considerations regarding obtaining signed consent to participate (i.e. whether she feels safe to conduct the interview over the phone if the perpetrator is in the household) (Downes et al., 2014).

In the consultation, one MBCP evaluator made the following statement regarding coercion:

One of the issues for me with evaluation over the years has been about getting the opinion from the family about how the participant is doing, and in particular men taking accountability and women's safety. So, I think it is uniquely tricky in MBCPs. You also don't know the level of coercion that goes on. So how can you be sure that she is not just saying that he has changed or improved? So [are] the data authentic or is the victim/survivor being pressured to say things? And if she does offer an authentic account, is there going to be retaliation?

Given the multitude of issues that could potentially put a victim/survivor at risk, evaluators need to work with women's specialist support workers to ensure that they are appropriately briefed about expectations of engagement in the evaluation and that necessary safeguards are in place to mitigate risk of further abuse and violence. In the first instance, victims/survivors should be approached by support workers to see if they consent to their contact information being shared with the evaluators.

Importantly, victims/survivors should be given an opportunity to make decisions about when and how they would like to be contacted and receive information about the evaluation. Specialist workers engage in various forms of communication with victims/survivors, including text updates (to inform women of men's participation in the MBCP), phone support, and one-to-one contact with the women's support worker (Downes et al., 2014). Specialist women's workers are also crucial to evaluation activities with victims/survivors. They are able to appropriately advise evaluators how best to make contact, ensure voluntary consent to be involved, and minimise risks to safety.

Potentially re-traumatising women who do not have adequate support

Evaluators need to consider the risks for victims/survivors who engage in the evaluation. Victims/survivors may need to be assessed in relation to their *current* situation and vulnerability to take part, with recognition that their capacity and agency may vary greatly over the course of the evaluation. As a result, women need to be approached as individuals, and the potential for harm needs to be viewed as being located on a continuum that may change over time.

In the context of a qualitative interview, retelling and revisiting traumatic events and circumstances may have unanticipated consequences for victims/survivors. Adequate debriefing following an interview is important to ensure they have necessary supports in place. However, given the possibility that distress triggered in the interview may also be delayed for some victims/survivors, there is considerable evidence to suggest that victims/survivors should only be involved in the evaluation if they are currently receiving assistance from a support worker.

Risk of disempowering victims/survivors

Victims/survivors need to be informed that their participation in the evaluation is voluntary, and they should not be unduly pressured to take part. Evaluators need to work closely with women's support workers to ensure that victims/survivors adequately understand their role in the evaluation and can make an autonomous choice to participate.

One evaluator we interviewed for this guide emphasised that often the interviews are an opportunity for women to tell their story. Many victims/survivors positioned themselves as active agents in the evaluation and were open to being interviewed. The evaluation needs to consider the extent to which the interview process empowers victims/survivors. Given the context, where women often have little control over their lives due to perpetrators' use of violence and controlling behaviour, providing them with an opportunity to discuss their experience and to offer feedback about how the program might need to change is an opportunity for them to exert some agency.

Careful management and collection of data from women must be undertaken to ensure that evaluation is not at risk of disempowering victims/survivors, intruding on their lives and choices, misrepresenting their experiences or assuming they are incapable of making independent decisions (McDermott & Garofalo, 2004).

Equally importantly, evaluators are encouraged to view victims/survivors as active and autonomous agents. Adopting ethical practice that attends to women’s potential empowerment in the evaluation process, often through the “telling of their story”, is important. The design of data collection instruments—such as the interview schedule—needs to be carefully considered to ensure victims/survivors are also provided opportunities to reflect and experience positive impacts.

Ethical approval and working with vulnerable groups

Any evaluation involving “vulnerable” groups will require ethics approval from a human research ethics committee (HREC). These groups include perpetrators, victims/survivors, and their children. If you are working with an external, experienced evaluator, such as a university, this organisation will obtain its own ethical approval. However, if you are conducting the evaluation yourself, you will need to do so. Some community service organisations have their own ethics approval processes, and there are some organisations that accept external ethics applications. The NHMRC keeps a list of registered ethics committees by state.⁷

There are some types of evaluation that will not require ethics approval because the activity is considered to have “negligible risk” associated with it. These evaluations generally include “quality assurance/audit projects that do not involve access to or collection of private, sensitive or health data” (University of Melbourne, n.d.). This might include activities such as collecting participant feedback on group sessions. However, if you are uncertain whether ethics approval is needed, it is best to speak to a member of a human research ethics committee to clarify. “Secondary” use of data (i.e. analysing existing data) may still require ethics approval even though it involves no direct human contact. Whether ethics approval is required will depend on factors such as whether any individual can be identified within the data (e.g. participants in programs), and whether the data are being used for a purpose outside that to which the person providing the data has consented (Tripathy, 2013). Again, it is best to clarify these issues with an ethics committee to ensure that any evaluation activities are conducted ethically.



⁷ The list of registered ethics committees can be found at <https://www.nhmrc.gov.au/research-policy/ethics/human-research-ethics-committees>

Further reading

Evaluation readiness

Davies, R. (2015). *Evaluability assessment*. Retrieved 21 February, 2020 from http://betterevaluation.org/themes/evaluability_assessment

Day, A., Vlasis, R., Chung, D., & Green, D. J. (2019). Evaluation readiness, program quality and outcomes in men's behaviour change programs (Research report, 01/2019). Sydney: ANROWS.

Kaufman-Levy, D., & Poulin, M. (2003). *Evaluability assessment: Examining the readiness for a program evaluation*. Retrieved from <http://www.jrsa.org/pubs/juv-justice/evaluability-assessment.pdf>

Evaluation planning

Australia's National Research Organisation for Women's Safety. (2020). *The views of Australian judicial officers on domestic and family violence perpetrator interventions* (Research to policy and practice, 13/2020). Sydney: ANROWS.

Better Evaluation. (2016a). Decide purpose. Retrieved from https://www.betterevaluation.org/en/rainbow_framework/frame/decide_purpose

Better Evaluation. (2016b). Specify the key evaluation questions (KEQs). Retrieved from https://www.betterevaluation.org/en/rainbow_framework/frame/specify_key_evaluation_questions

CoastAdapt. (2017). Identifying indicators for monitoring and evaluation. Retrieved from <https://coastadapt.com.au/how-to-pages/identifying-indicators-monitoring-and-evaluation>

International Development Research Centre. (2012). *Identifying the intended user(s) and use(s) of an evaluation*. Retrieved from <https://www.betterevaluation.org/sites/default/files/idrc.pdf>

Transferability and replicability of evaluation findings

Catch-22. (n.d.). *Evidence is confidence: How to create a richer evidentiary tapestry*. Retrieved from <https://cdn.catch-22.org.uk/wp-content/uploads/2015/11/Realising-Ambition-Programme-Insight-Evidence-is-Confidence.pdf>

Wang, S., Moss, J. R., & Hiller, J. E. (2006). Applicability and transferability of interventions in evidence-based public health. *Health Promotion International*, 21(1), 76–83. <https://doi.org/10.1093/heapro/dai025>

Most Significant Change technique

Davies, R., & Dart, J. (2005). *The “most significant change” (MSC) technique: A guide to its use*. Retrieved from https://www.researchgate.net/publication/275409002_The_'Most_Significant_Change'_MSC_Technique_A_Guide_to_Its_Use

Collective impact

Australian Institute of Family Studies. (2017). *Collective impact: Evidence and implications for practice*. Retrieved from <https://aifs.gov.au/cfca/publications/collective-impact-evidence-and-implications-practice/evidence-and-evaluation>

Cairns Regional Domestic Violence Service. (n.d.). Cairns collective impact on domestic and family violence. Retrieved from <https://www.dvcairns.org/collective-impact-on-dv>

Cardinia. (n.d.). Preventing family violence: Together we can. Retrieved 21 February, 2020 from <http://togetherwecan.org.au/>

Logan Together. (2017). Resources. Retrieved from <https://logantgether.org.au/resources/>

Internal versus external evaluators

Conley-Tyler, M. (2005). A fundamental choice: Internal or external evaluation? *Evaluation Journal of Australasia*, 4(1–2), 3–11. <https://doi.org/10.1177%2F1035719X05004001-202>

Ethics

Australasian Evaluation Society. (2013). *Guidelines for the ethical conduct of evaluations*. Retrieved from https://www.aes.asn.au/images/stories/files/membership/AES_Guidelines_web_v2.pdf

National Health and Medical Research Council, Australian Research Council, & Universities Australia. (2018). *National statement on ethical conduct in human research 2007* (Rev. ed.). Canberra: NHMRC.

The National Health and Medical Research Council list of registered ethics committees by state is available at <https://www.nhmrc.gov.au/research-policy/ethics/human-research-ethics-committees>

Evaluation examples

Blagg, H., Williams, E., Cummings, E., Hovane, V., Torres, M., & Woodley, K. N. (2018). *Innovative models in addressing violence against Indigenous women: Final report* (ANROWS Horizons, 01/2018). Sydney: ANROWS.

Project information available at: <https://www.anrows.org.au/project/evaluation-of-innovative-models-of-interagency-partnerships-collaboration-coordination-and-or-integrated-responses-to-family-and-or-sexual-violence-against-women-in-australian-indigenous-communities/>

Quadara, A., O'Brien, W., Ball, O., Douglas, W., & Vu, L. (2020). *Good practice in delivering and evaluating interventions for young people with harmful sexual behaviours* (Research report, 17/20). Sydney: ANROWS.

Project information available at: <https://www.anrows.org.au/project/good-practice-in-delivering-and-evaluating-interventions-for-young-people-with-harmful-sexual-behaviours/>

References

- Arai, L., Heawood, A., Feder, G., Howarth, E., MacMillan, H., Moore, T. H., . . . Gregory, A. (2019). Hope, agency, and the lived experience of violence: A qualitative systematic review of children's perspectives on domestic violence and abuse. *Trauma, Violence, & Abuse*, 20(3), 1–12. <https://doi.org/10.1177/1524838019849582>
- Australian Institute of Family Studies. (2017). *Collective impact: Evidence and implications for practice*. Retrieved 21 February, 2020 from <https://aifs.gov.au/cfca/publications/collective-impact-evidence-and-implications-practice/evidence-and-evaluation>
- Australasian Evaluation Society. (2013). *Guidelines for the ethical conduct of evaluations*. Retrieved from https://www.aes.asn.au/images/stories/files/membership/AES_Guidelines_web_v2.pdf
- Bender, A. K. (2017). Ethics, methods, and measures in intimate partner violence research: The current state of the field. *Violence Against Women*, 23(11), 1382–1413. <https://doi.org/10.1177/1077801216658977>
- Better Evaluation. (2016a). Decide purpose. Retrieved from https://www.betterevaluation.org/en/rainbow-framework/frame/decide_purpose
- Better Evaluation. (2016b). Specify the key evaluation questions (KEQs). Retrieved from https://www.betterevaluation.org/en/rainbow-framework/frame/specify_key_evaluation_questions
- Brown, T., Flynn, C., Fernandez Arias, P., & Clavijo, C. (2016). *A study of the impact on men and their partners in the short term and in the long term of attending men's behaviour change programs*. Retrieved from <https://violencefreefamilies.org.au/longitudinal-study-effectiveness-mens-behaviour-change-programs/>
- Cairns Regional Domestic Violence Service. (n.d.). Cairns collective impact on domestic and family violence. Retrieved 21 February, 2020 from <https://www.dvcairns.org/collective-impact-on-dv>
- Cardinia. (n.d.). Preventing family violence: Together we can. Retrieved 21 February, 2020 from <http://togetherwecan.org.au/>
- Catch-22. (n.d.). *Evidence is confidence: How to create a richer evidentiary tapestry*. Retrieved from <https://cdn.catch-22.org.uk/wp-content/uploads/2015/11/Realising-Ambition-Programme-Insight-Evidence-is-Confidence.pdf>
- CoastAdapt. (2017). *Identifying indicators for monitoring and evaluation*. Retrieved from <https://coastadapt.com.au/how-to-pages/identifying-indicators-monitoring-and-evaluation>
- Conley-Tyler, M. (2005). A fundamental choice: Internal or external evaluation? *Evaluation Journal of Australasia*, 4(1–2), 3–11. <https://doi.org/10.1177%2F1035719X05004001-202>
- Davies, R. (2015). *Evaluability assessment*. Retrieved 21 February, 2020 from http://betterevaluation.org/themes/evaluability_assessment
- Davies, R., & Dart, J. (2005). *The "most significant change" (MSC) technique: A guide to its use*. Retrieved from https://www.researchgate.net/publication/275409002_The_'Most_Significant_Change'_MSC_Technique_A_Guide_to_Its_Use
- Day, A., & Casey, S. (2010). Maintaining change programmes for offenders: Some suggestions for practice. *Psychology, Crime & Law*, 16(6), 449–458. <https://doi.org/10.1080/10683160902798086>
- Day, A., Vlasis, R., Chung, D., & Green, D. J. (2019). *Evaluation readiness, program quality and outcomes in men's behaviour change programs* (Research report, 01/2019). Sydney: ANROWS.

- Diemer, K., Humphreys, C., Fogden, L., Gallant, D., Spiteri-Staines, A., Bornemisza, A., & Vercoe, E. (2020). *Caring Dads program: Helping fathers value their children: Three site independent evaluation 2017–2019* (Final report). Retrieved from https://mcusercontent.com/735c980d83254d8db51ef91e3/files/294308b1-1756-43c2-88d4-2a2eac8e4e05/2020_Caring_Dads_Final_Report_3_3_2020.pdf?mc_cid=979b3fcc6e&mc_eid=1737a9d628
- Downes, J., Kelly, L., & Westmarland, N. (2014). Ethics in violence and abuse research: A positive empowerment approach. *Sociological Research Online*, 19(1), 1–13. <https://doi.org/10.5153/sro.3140>
- Dutton, D. G., & Corvo, K. (2006). Transforming a flawed policy: A call to revive psychology and science in domestic violence research and practice. *Aggression and Violent Behavior*, 11(5), 457–483. <https://doi.org/10.1016/j.avb.2006.01.007>
- Dutton, M. A., Holtzworth-Munroe, A., Jouriles, E., McDonald, R., Krishnan, S., McFarlane, J., & Sullivan, C. (2003). *Recruitment and retention in intimate partner violence research*. Rockville, MD: US Department of Justice.
- Gondolf, E. W. (2012). *The future of batterer programs: Reassessing evidence-based practice*. Boston: Northeastern University Press.
- Hawe, P., Degeling, D., & Hall, J. (1990). Evaluability assessment: Getting ready to test programme effects. In P. Hawe, D. Degeling, & J. Hall (Eds.), *Evaluating health promotion: A health worker's guide* (pp. 86–100). Artarmon, NSW: MacLennan & Petty.
- Hearn, J., Andersson, K., & Cowburn, M. (2007). *Background paper on guidelines for researchers on doing research with perpetrators of sexual violence*. Pretoria: Sexual Violence Research Initiative.
- Howarth, E., Moore, T. H., Stanley, N., MacMillan, H. L., Feder, G., & Shaw, A. (2019). Towards an ecological understanding of readiness to engage with interventions for children exposed to domestic violence and abuse: Systematic review and qualitative synthesis of perspectives of children, parents and practitioners. *Health & Social Care in the Community*, 27(2), 271–292. <https://doi.org/10.1111/hsc.12587>
- International Development Research Centre. (2012). *Identifying the intended user(s) and use(s) of an evaluation*. Retrieved from <https://www.betterevaluation.org/sites/default/files/idrc.pdf>
- Kania, J., & Kramer, M. (2011). Collective impact. *Stanford Social Innovation Review*, 9(1), 36–41. Retrieved from https://ssir.org/articles/entry/collective_impact
- Kaufman-Levy, D., & Poulin, M. (2003). *Evaluability assessment: Examining the readiness for a program evaluation*. Retrieved from <http://www.jrsa.org/pubs/juv-justice/evaluability-assessment.pdf>
- Kellogg, W. K. (2004). *Using logic models to bring together planning, evaluation, and action: Logic model development guide*. Michigan: W.K. Kellogg Foundation.
- Kelly, L., & Westmarland, N. (2015). *Domestic violence perpetrator programmes: Steps towards change—Project Mirabal final report*. London and Durham: London Metropolitan University and Durham University.
- Knowlton, L. W., & Phillips, C. C. (2012). *The logic model guidebook: Better strategies for great results* (2nd ed.). London: Sage Publishing.
- Lilley-Walker, S. J., Hester, M., & Turner, W. (2018). Evaluation of European domestic violence perpetrator programmes: Toward a model for designing and reporting evaluations related to perpetrator treatment interventions. *International Journal of Offender Therapy and Comparative Criminology*, 62(4), 868–884. <https://doi.org/10.1177/0306624X16673853>

- Logan, T. K., Walker, R., Shannon, L., & Cole, J. (2008). Combining ethical considerations with recruitment and follow-up strategies for partner violence victimization research. *Violence Against Women, 14*(11), 1226–1251. <https://doi.org/10.1177/1077801208323976>
- Mackay, E., Gibson, A., Lam, H., & Beecham, D. (2015). *Perpetrator interventions in Australia: Part One—Literature review* (State of knowledge paper) (ANROWS Landscapes, 01/2015). Sydney: ANROWS.
- Mark, M., Henry, G., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco: Jossey-Bass Inc.
- McDermott, M. J., & Garofalo, J. (2004). When advocacy for domestic violence victims backfires: Types and sources of victim disempowerment. *Violence Against Women, 10*(11), 1245–1266. <https://doi.org/10.1177/1077801204268999>
- McKenzie, J., White, T., Minty, R., & Clancy, L. (2016). *Program planning and evaluation: How to use evidence to support communities for children services*. Melbourne: Centre for Family Research and Evaluation.
- McLaren, H., Fischer, J., & Zannettino, L. (2020). *Defining quality of life indicators for measuring perpetrator intervention effectiveness* (Research report, 05/2020). Sydney: ANROWS
- National Health and Medical Research Council, Australian Research Council, & Universities Australia. (2018). *National statement on ethical conduct in human research 2007* (Rev. ed.). Canberra: NHMRC.
- No to Violence. (2006). *Men's behaviour change group work: Minimum standards and quality practice*. Melbourne: NTV.
- Opitz, C. (2014). Considerations for partner contact during men's behaviour change programs: Systemic responses and engagement. *Ending Men's Violence Against Women and Children: The No To Violence Journal, Autumn*, 114–142. Retrieved from <http://shantiworks.com.au/wp-content/uploads/2014/08/2014-ntv-journal-tc-article.pdf>
- Patton, M. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks: Sage Publishing.
- Paymar, M., & Barnes, G. (2007). *Countering confusion about the Duluth model*. Duluth, MN: Domestic Abuse Intervention Programs.
- Respect. (2017). *Respect outcomes framework*. Retrieved from <https://www.escb.co.uk/media/1631/respect-outcomes-framework-15-11-17.pdf>
- Rychetnik, L., Frommer, M., Hawe, P., & Shiell, A. (2002). Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology & Community Health, 56*(2), 119–127. <https://doi.org/10.1136/jech.56.2.119>
- Salabarría-Peña, Y., Apt, B. S., & Walsh, C. M. (2007). *Practical use of program evaluation among sexually transmitted disease (STD) programs: Vol. 4—Gather credible evidence*. Atlanta: Centers for Disease Control and Prevention.
- Semiatin, J. N., Murphy, C. M., & Elliott, J. D. (2013). Observed behavior during group treatment for partner-violent men: Acceptance of responsibility and promotion of change. *Psychology of Violence, 3*(2), 126–139. <https://doi.org/10.1037/a0029846>
- Silvergleid, C. S., & Mankowski, E. S. (2006). How batterer intervention programs work: Participant and facilitator accounts of processes of change. *Journal of Interpersonal Violence, 21*(1), 139–159. <https://doi.org/10.1177/0886260505282103>
- Tharp, A. T., DeGue, S., Lang, K., Valle, L. A., Massetti, G., Holt, M., & Matjasko, J. (2011). Commentary on Foubert, Godin, & Tatum (2010): The evolution of sexual violence prevention and the urgency for effectiveness. *Journal of Interpersonal Violence, 26*(16), 3383–3392. <https://doi.org/10.1177/0886260510393010>
- Tripathy, J. P. (2013). Secondary data analysis: Ethical issues and challenges. *Iranian Journal of Public Health, 42*(2), 1478–1479. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4441947/>

University of Melbourne (n.d.). Do I need ethics approval? Retrieved from <https://staff.unimelb.edu.au/research/ethics-integrity/human-ethics/faq/do-i-need-ethics-approval>

Vlais, R., Campbell, E., & Green, D. (2019). *Foundations for family and domestic violence perpetrator interventions systems*. Melbourne: Centre for Innovative Justice, RMIT University.

Vlais, R., & Green, D. (2018). *Developing an outcomes framework for men's behaviour change programs*. Perth: Stopping Family Violence.

Vlais, R., Ridley, S., Green, D., & Chung, D. (2017). *Family and domestic violence perpetrator programs: Issues paper of current and emerging trends, developments and expectations*. Perth: Stopping Family Violence.

Walby, S., Towers, J., Balderston, S., Corradi, C., Francis, B. J., Heiskanen, M., . . . Strid, S. (2017). *The concept and measurement of violence against women and men*. Bristol: Policy Press.

Wang, S., Moss, J. R., & Hiller, J. E. (2006). Applicability and transferability of interventions in evidence-based public health. *Health Promotion International*, 21(1), 76–83. <https://doi.org/10.1093/heapro/dai025>

Westmarland, N., Kelly, L., & Chalder-Mills, J. (2010). *Domestic violence perpetrator programmes: What counts as success?* London: London Metropolitan University and Durham University.

Wistow, R., Kelly, L., & Westmarland, N. (2017). "Time out": A strategy for reducing men's violence against women in relationships? *Violence Against Women*, 23(6), 730–748. <https://doi.org/10.1177/1077801216647944>

Work with Perpetrators of Domestic Violence–European Network (WWP–EN). *Project IMPACT*. Retrieved 21 February, 2020 from <https://www.work-with-perpetrators.eu/projects/project-impact>



ANROWS

AUSTRALIA'S NATIONAL RESEARCH
ORGANISATION FOR WOMEN'S SAFETY

to Reduce Violence against Women & their Children