



**IMPROVING
EFFORTS TO
PREVENT
CHILDREN'S
EXPOSURE
TO VIOLENCE**



**A handbook to support
the evaluation of child
maltreatment prevention
programmes**

Improving efforts to prevent children's exposure to violence

A HANDBOOK FOR DEFINING
PROGRAMME THEORY AND PLANNING
FOR EVALUATION IN THE NEW
EVIDENCE-BASED CULTURE



WHO Library Cataloguing-in-Publication Data :

Improving efforts to prevent children's exposure to violence:
a handbook for defining program theory and planning for
evaluation in the new evidence-based culture.

1.Violence – prevention and control. 2.Child. 3.Stress Disorders,
Traumatic. 4.Child Abuse. 5.Program Development. I.Jones,
Lisa M. II.World Health Organization. III.University of New
Hampshire. Crimes against Children Research Center.

ISBN 978 92 4 150788 2

(NLM classification: WA 325)

© **World Health Organization 2014**

All rights reserved. Publications of the World Health Organization are available on the WHO website (www.who.int) or can be purchased from WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland (tel.: +41 22 791 3264; fax: +41 22 791 4857; e-mail: bookorders@who.int). Requests for permission to reproduce or translate WHO publications – whether for sale or for non-commercial distribution – should be addressed to WHO Press through the WHO website (www.who.int/about/licensing/copyright_form/en/index.html).

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by the World Health Organization to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall the World Health Organization be liable for damages arising from its use.

Designed by minimum graphics

Printed in Luxembourg

Contents

Acknowledgements	iv
Preface	v
Introduction	1
Purpose of the handbook	1
Structure and content of the handbook	2
1. Why research and evaluation is critical to preventing children's exposure to violence	3
2. Partnering with research and evaluation specialists	9
Collaborating with outside consultants or research partners	9
Finding an evaluation partner	11
Working with your evaluation partners	11
Budgeting for evaluation	12
Ethical issues	12
3. Defining your programme theory	14
Programme theory: the basic concepts	14
Extending your programme theory model to include causal and risk factors	17
Finding the research on your programme theory elements	21
4. Choosing and measuring outcomes	22
Defining outcomes: common pitfalls	22
Outcome measurement strategies and tools	24
Additional considerations	26
5. Understanding programme evaluation	28
What does it mean to have an evidence-based programme?	28
Preparing for evaluation	29
Different kinds of evaluation	29
Types of outcome evaluation	30
6. Applying this handbook to your programmes	34
References	35
Appendix: Evaluation resources	39

Acknowledgements

This report was written by Lisa M. Jones, Ph.D., Research Associate Professor of Psychology at the Crimes Against Children Research Center (CCRC), University of New Hampshire (UNH). Special thanks to Dr. Chris Mikton at the World Health Organization (WHO), Dr. Lucie Cluver at the Department of Social Policy & Intervention, Oxford University and the Department of Psychiatry & Mental Health, University of Cape Town, Dr. David Finkelhor at the Crimes Against Children Research Center, UNH and Patricia Lannen at UBS Optimus Foundation for critical feedback on earlier versions of this report. Funds for completion of this report were provided by UBS Optimus Foundation. The report was produced by WHO and edited by Ann Morgan.

Cite as: Jones L. *Improving efforts to prevent children's exposure to violence: A handbook for defining program theory and planning for evaluation in the evidence-based culture*. Geneva, Switzerland, World Health Organization, 2014.

Introduction

As part of a global movement to direct greater attention and resources to child protection, programmes aimed at reducing children's exposure to violence are being implemented with increasing frequency across the world. These programmes are diverse and range from raising public awareness of the issues to widening and strengthening government policies and protective structures, improving children's and families' access to medical, therapeutic and legal support, and increasing children's and parents' protective skills.

Over the last several decades a substantial body of research has accumulated, providing much needed information about the experience of violence in childhood. We now know that child abuse, neglect, peer violence, sexual victimization and exposure to domestic and community violence are suffered by very large numbers of children. We understand much more about the serious negative consequences of violence exposure on children's physical and mental health, effects that can carry through from generation to generation. In addition, we have come to appreciate that children's health and safety concerns often cluster together, and that children and families with the fewest resources often suffer from disproportionate levels of exposure to violence.

Evaluation research has also helped us to identify which prevention and intervention strategies work well and make the biggest difference to children's well-being. What is needed now is a global movement to build on this knowledge in order to channel programme efforts in increasingly effective ways. Given the seriousness of the problem of children's exposure to violence and the scarcity of resources with which to tackle it, our focus should be to ensure that:

- new programmes and initiatives are grounded in research knowledge;
- programmes are routinely evaluated so that we continue to learn more about what works and what does not.

It is only by sharing, using and growing the evidence base that we will hasten our positive impact on the lives and well-being of children worldwide.

Purpose of the handbook

This handbook is intended to help implementing agencies (e.g. nongovernmental organizations (NGOs), development/foreign aid agencies, community-based organizations, not-for-profit agencies) make better use of existing research and plan for evaluation when designing and implementing child violence prevention programmes, and also to convey these intentions to potential funding organizations.

In recent years, those who are funding programme work in child protection have come to recognize the importance of identifying and building on proven solutions. They are increasingly

asking applicants to explain how their proposed programme is informed by current research and to outline their plans to evaluate the effectiveness of their programme. Funding bodies are also aware that some evaluation study designs are better – more robust and therefore more informative – than others, and that it can be difficult for programme implementers, who typically do not have experience in research and evaluation, to meet all of these expectations by themselves. For this reason, there is a trend among funders to look more favourably on agencies that partner with researchers to help design research-based interventions and to assist with evaluation studies.

Thus, the principal aim of this handbook is to serve as a reference for agencies seeking to improve their ability to meet the current expectations of funding agencies. While they are primarily written for those who implement programmes aimed at reducing children's exposure to violence, the information contained herein is applicable to a wide range of social problem interventions.

Structure and content of the handbook

The handbook opens with a discussion of the reasons why evidence and research are considered so important in the context of advancing efforts to improve children's safety and reduce their exposure to violence. In **SECTION 1: *Why research and evaluation is critical to preventing children's exposure to violence***, we address common questions and concerns that many people have about the need for evaluation: Why spend time and money on evaluation? Can it really help us improve children's safety and well-being? Does it provide implementing agencies with useable information?

In **SECTION 2** of the handbook, *Partnering with research and evaluation specialists*, we explain how research advisors and partners can support and guide agencies through the process of designing a programme and applying for funding; we also offer advice on how to go about establishing such a partnership.

SECTION 3: *Defining your programme theory* provides step-by-step guidance, illustrated with examples, on formulating a research-based programme theory model and explains why developing such a model is so important. It also outlines the advantages of working with research partners who can help you identify relevant research to support your programme theory linkages.

In **SECTION 4: *Choosing and measuring outcomes*** we guide you through the process of choosing appropriate programme outcomes – information that you can use to judge progress towards your programme's stated aims and objectives. Defining specific and measurable programme outcomes and choosing sound measurement strategies is an important prerequisite to a strong evaluation.

SECTION 5: *Understanding programme evaluation* explains, in non-specialist language, what distinguishes a good evaluation study (i.e. one that meets basic methodological standards) from other types of programme evaluation or data collection. The intention of this section is to provide enough information to enable programme developers to have an informed discussion about their evaluation needs with research and evaluation partners.

In the last section, **SECTION 6: *Applying the handbook to your programmes*** we provide summary suggestions on using the handbook to increase the use of research and evaluation to support your programmes and interventions.

Why research and evaluation is critical to preventing children's exposure to violence

Programme developers sometimes say that contributing to research and evaluation feels like a luxury that they cannot afford. Research and evaluation require time and money, commodities that are often in short supply. Those who observe the impact of their programme on families on a daily basis may question the need for an evaluation, thinking that the observed impact should represent sufficient evidence of a programme's success. In this section, we address these and other common concerns about evaluation.

1. I don't have time for evaluation

One frequent complaint about conducting evaluations is that the process is too slow – it takes too long to see results – and programme developers have expressed concern that evaluation impedes implementation. To many developers, who are acutely aware of the serious harm that violence and abuse can do to children, the price of slowing down implementation efforts feels too high.

We would argue that, in the longer term at least, the social costs of implementing an intervention without the support of an evaluation process are also potentially very high. Not evaluating could mean that money is wasted on ineffective programmes; that communities and policy-makers become frustrated or disillusioned; or that there are lengthy delays in providing families and children with effective services and programmes. There is also the very real possibility that untested programmes could make things worse in unexpected ways, via the so-called “boomerang effect” (see **BOX 1.1**). Those working in the field of child protec-

BOX 1.1: THE BOOMERANG EFFECT

■ One of the things that can happen when you ignore past research and forego evaluation is that on occasion, a programme – even one designed and implemented with best of intentions – can have an effect opposite to that which was originally intended. This is called a “boomerang effect”.

Efforts to steer youth away from drug and alcohol use in the United States of America provide a classic case in point. Several public health researchers have established that issuing warnings about the dangers of using drugs, cigarettes or alcohol sometimes made these substances more attractive to young people and actually encouraged their consumption (See, for example, Werch & Owen, 2002). Boomerang effects have also been observed following the introduction of interventions aimed at dissuading people from eating unhealthy foods (Ringold, 2002), and fear-based juvenile delinquency rehabilitation programmes such as Scared Straight (Petrosino, Turpin-Petrosino & Buehler, 2003)*.

Several reasons have been proposed to explain why boomerang effects can happen. Some people bridle at being told what to do and adopt oppositional attitudes and behaviours. Others simply may wish to try something once it has been mentioned to them, even if the behaviour is risky or harmful (Petrosino, Turpin-Petrosino & Buehler, 2003).

tion must, like doctors, consider themselves bound to “first do no harm”. The presence of boomerang effects means that untested programmes cannot be assumed to be harmless. The assumption “even if it doesn't work perfectly, it won't make things worse” may in fact be wrong. This is one of the reasons why it is so important to evaluate the effects of interventions, especially on vulnerable populations such as children.

Devoting time and resources for evaluation is also likely to reap direct programme benefits. If you can provide evidence that your programme works, your chances of securing continued funding, thereby ensuring your programme's sustainability, will be that much greater.

2. I don't need an evaluation to know my programme works

Programme developers and agencies will sometimes claim that they “know” their programme is helping people and that evaluation is not needed. Such claims may be based on seeing some families and children do better and on the gratitude expressed by people in the community.

Unfortunately, evaluation research has shown that people's intuitive judgements about programmes can often be wrong (Farrelly, Niederdeppe & Yarsevich, 2003; Finkelhor, 1979; Petrosino, Turpin-Petrosino & Buehler, 2003; Ringold, 2002). Here are a few experiences which may suggest to some stakeholders that a programme is performing well – and the reasons why such claims might be deceptive and unreliable as evidence:

↳ **People react very positively to my programme**

Positive reactions do not necessarily translate into lasting prevention and treatment effects for a number of reasons. For instance, people can like a programme and respond positively to it simply because they like the people involved, not because it provides any lasting benefit to them. To those who are suffering or are in great need, any kind of attention or help may elicit a positive reaction, making the impact of a programme appear more significant than it actually is. Also, community members may be reluctant to give negative feedback through fear of a loss of resources or because of power disparities.

↳ **I have specific examples of how my programme helped a family**

Anecdotes are compelling but by themselves are not adequate evidence that a programme is effective. Anecdotal evidence is often subjective and prone to bias. Furthermore, it is not enough to show that a handful of children or families were helped. A successful programme should be able to demonstrate that it has improved outcomes for many children and their families.

↳ **I know from my experience and professional background that this programme is a good one**

Your programme may indeed be having the desired impact. However, the only way to be sure, and to convince others of its effectiveness, is to conduct an evaluation. We all have a strong inclination to want our efforts to succeed and may discount or be blind to information that suggests otherwise. It is critical therefore that programme developers and implementers are aware of the likelihood of preconceptions, biases and blind-spots as far as the impact of their programme is concerned.

3. I'm worried the results will not show that my programme is effective

One reason why people may avoid evaluation is a fear of negative results. In addition, programme developers may be concerned that an evaluator will not reliably measure all of

the ways in which their programme could be valuable to its recipients and thus come to an erroneous conclusion that the entire programme is failing outright.

In practice, the results of an evaluation are rarely wholly negative or wholly positive. Some of the best evaluations conducted in this field have shown that programmes are achieving positive results in terms of some outcomes but not in others, or are working for some sub-populations but not for others. Furthermore, negative results do not necessarily mean failure; there are important lessons that can be learned from “null” results.

An honest appraisal is always going to be a valuable and worthwhile exercise. The very process of preparing for an evaluation will help you to clarify which of your interventions you think will be most effective and to define how and why you think this to be the case. In addition, whatever they are, evaluation results provide much useful information that can be fed back into, and guide, future programme development. Even when there is little to show in the way of success, programme developers who have engaged in evaluation are still able to publicize what they have learned, and how they are modifying their programme as a result. In the current culture of respect for evidence, a well-evaluated programme still comes out ahead.

4. I don't have the background to do an evaluation

While the majority of agency professionals working in the field of child protection have an appreciation of the value of evidence-based practice and the need to assess programme effectiveness, many do not possess the necessary expertise to carry out an evaluation. It takes specialist skills to identify and summarize academic research on a particular topic and very specific skill sets to design an experimental study to evaluate the impact of a given intervention.

Although one of the primary aims of this handbook is to demystify the research and evaluation process, it is recognized that this type of work requires skills that many programme developers or administrators simply do not have. For this reason, we strongly encourage programme agencies and developers to collaborate with researchers, university staff and evaluators who do have the relevant background and experience in evaluation study design. Suggestions on how to connect with academics and researchers are provided in the next section of the handbook.

5. Has evaluation research really advanced our understanding of what works best to protect children and improve their well-being?

Programme developers may be a bit sceptical about the long-term benefits of evaluation research beyond supporting the work of an individual programme. However, there are numerous examples of how evaluation research has advanced knowledge on a given approach to violence prevention or treatment (Chaffin & Friedrich, 2004; MacMillan, Wathen, Barlow, Fergusson et al., 2009). A select list is given below:

- *Parenting group education programmes.* Parenting group education programmes are designed to help improve parenting skills and in turn to reduce mental health, emotional and behavioural problems in children. Examples include the Incredible Years programme¹ and the “Triple P” interventions.² The effectiveness of the Incredible Years programme,

¹ For further information, see: <http://www.incredibleyears.com/>.

² For further information, see: <http://www.triplep.net/>.

a US-based initiative targeting families with children aged 0–12 years which has subsequently been adopted in more than 20 countries including a number of low- and middle-income countries, has been confirmed in several randomized controlled trials (RCTs³) (Hutchings, Bywater, Eames & Martin, 2008, Baker-Henningham, Scott, Jones & Walker, 2012). Similarly, evaluations of the popular Triple P programme – the Positive Parenting Programme – have generally been positive, reporting significant improvements in child behaviour among participating families compared with controls (Nowak & Heinrichs, 2008; Sanders, Turner & Markie-Dadds, 2002). Building on these findings, parent education researchers are now working on identifying the key elements of good parenting programmes, improving their translation to low- and middle-income communities (Knerr, Gardner & Cluver, 2013) and determining how to increase their efficacy with respect to outcomes like child maltreatment.

- ↳ *Home visiting programmes.* Home visiting programmes are offered to pregnant women and/or to parents of infants and young children and are aimed at improving a range of parental and child health indicators, including the prevalence of child abuse. While evaluation findings have been mixed (MacMillan, Wathen, Barlow, Fergusson et al., 2009), there is some evidence that home visiting programmes do help, under some circumstances at least, to reduce the number of child protection reports and injuries and the frequency of self-reported abuse by mothers. Efforts are now under way to identify which elements of home visiting programmes are the most effective in terms of outcomes (Filene, Kaminski, Valle & Cachat, 2013), and to increase the number of evaluations of adaptations to the basic model in low- and middle-income countries (Rahman, Iqbal, Roberts & Husain, 2009; Wessels, Mikton, Ward, Kilbane et al., 2013).
- ↳ *Prevention programmes targeting peer violence.* Several countries have implemented education and training programmes, mainly school-based, aimed at reducing the problems associated with peer-to-peer violence. Meta-analyses of social–emotional learning (SEL) programmes provided in and out of schools have provided evidence of a generally positive impact of such programmes on a range of child well-being indicators (Durlak, Weissberg, Dymnicki, Taylor et al, 2011; Durlak, Weissberg & Pachan, 2010). Results of studies which assess the effectiveness of bullying prevention programmes have been more variable (Swearer, Espelage, Vaillancourt & Hymel, 2010), but as the number of studies increases and the data accumulate, researchers are improving their understanding of which peer violence prevention and intervention strategies work best.
- ↳ *Teen relationship violence prevention programmes.* Evidence in support of teen relationship violence prevention programmes is also accumulating. Several school-based programmes have been shown, in RCTs, to reduce levels of psychological, sexual and physical violence perpetrated against teens by their romantic partners (Foshee, Bauman, Arriaga, Helms et al., 1998; Wolfe, Crooks, Jaffe, Chiodo et al., 2009). The Centers for Disease Control (CDC) in the United States is using one of these interventions, Safe Dates, as the basis of a comprehensive violence prevention strategy called Dating Matters (Tharp, 2012).⁴ Evaluations of several other programmes and interventions aimed at preventing intimate partner violence in adolescents have provided evidence of their success (De Koker, Mathews, Zuch, Bastien et al., 2014), and efforts are ongoing to identify effective

3 A randomized controlled trial (RCT) is a highly rigorous evaluation methodology which is widely regarded as the “gold standard” when it comes to evaluation study design. An evaluation that employs a RCT methodology is generally considered to be superior to other designs in terms of the confidence that can be placed in its results. See Section 5 for more information.

4 For further details, see: <http://vetoviolence.cdc.gov/datingmatters/>.

relationship programmes that are appropriate for use in low- and middle-income countries (PREPARE, 2014).

- *Trauma-focused, cognitive-behavioural therapy.* Trauma-focused, cognitive-behavioural therapy (TF-CBT)⁵ is a systematic mental health treatment for traumatized children and their families. To date it has been tested in more than 10 evaluation studies, including several RCTs, which show that children who have received TF-CBT exhibit fewer traumatic symptoms and better mental health and coping than those who received alternate therapies, with lasting effects (Cohen, Deblinger, Mannarino & Steer, 2004).
- *Parent–child interaction therapy.* Parent–child interaction therapy (PCIT)⁶ is a treatment programme for child conduct disorders that is focused on improving the parent–child relationship and interaction patterns. Therapists provide guidance to parents in real time, often using communications technology or one-way mirrors. Its impact on physical abuse outcomes has been evaluated in several RCTs; results are generally positive and show that participants offered PCIT have a significantly reduced likelihood of a re-report of abuse compared with controls who were not (Chaffin, Silovsky, Funderburk, Valle et al., 2004; MacMillan, Wathen, Barlow, Fergusson et al., 2009). The use of PCIT in international settings is increasing, although work is still needed to understand how it can be used to best effect in low-resource settings and communities.

This list of examples serves to illustrate how evaluation research is improving our collective knowledge about how to best prevent children’s exposure to violence and its consequences. Note that in these examples evaluation is an ongoing process, building knowledge and understanding with each study extending the learning that came before. Through this process, we move towards implementing better and more effective solutions to the problems associated with children’s exposure to violence.

On the other hand, without evaluation, developers risk creating interventions that are based on faulty assumptions. This can lead to a waste of public resources should such a programme become popular and widely disseminated and then is later discovered to not have made much of a difference. Such was the case with drug abuse prevention programmes introduced in the United States in the 1970s and 1980s (see **BOX 1.2**).

Perhaps even worse, without evaluation, innovative and potentially beneficial programmes risk being overlooked. Consider a scenario in which a programme developer’s careful efforts result in a programme that does work but no outcomes are documented. Then, because effectiveness has not been demonstrated, the programme fails to obtain additional or sustained funding, eventually the efforts stop, and the chance for others to learn from, and build on this good work, is lost.

6. How do I make sure that my programme contributes to the growing evidence-base?

To contribute to the culture of evidence, agencies or professionals involved in programme development and implementation should plan to do four things:

1. Develop and build ongoing partnerships with experts in evaluation and research.
2. Specify programme theories that build on the current research knowledge. In practice, this means using interventions that target those risk and protective factors that have

5 For further information on these therapies, see: <http://tfcbt.musc.edu/> or <http://www.childwelfare.gov/pubs/trauma/trauma.pdf>.

6 For further information on this form of therapy, see <http://www.pcit.org/>.

BOX 1.2: WHAT CAN GO WRONG WHEN EVALUATION COMES TOO LATE: US DRUG ABUSE PREVENTION PROGRAMMES OF THE 1970S AND 1980S

■ Rising drug use among youth in the United States during the 1960s and 1970s prompted a slew of prevention education programmes to respond to what was seen as an emergency at that time. Many of these focused on delivering messages about the dire effects of drug use. The most popular of these programmes, DARE, was rapidly adopted by as many as 80% of school systems around the country, bringing law enforcement officials into schools to warn adolescents about the dangers of drug taking.

However, when researchers began to evaluate DARE – well after it had become the programme of choice – it was found to be largely ineffective. After reviewing six large-scale, long-term evaluations, a US General Accounting Office report in 2003 (GAO-03-172R), found “no significant differences in illicit drug use between students who received DARE... and ...students who did not” (Kanof, 2003). It took a second generation of programmes and evaluations before more effective approaches were developed and implemented. These newer approaches emphasize resistance skills and focus on changing peer group norms.

Unfortunately, 20 years, millions of dollars and hours of education time were wasted on ineffective prevention. However, two key lessons were learned. One: programmes that excite people and that “feel” right are not necessarily the ones that work best, and two: once programmes become entrenched, even ineffective ones, they become very hard to replace.

been proven to affect your chosen outcomes. It also means implementing or adapting existing evidence-based interventions as far as possible.

3. Define specific, valid and measurable outcomes.
4. Incorporate evaluation plans as you design and seek funding for your programme.

Increasingly, foundations and other funders are requiring these elements to be outlined in funding applications. The sections that follow are designed with these aims in mind and provide more detailed guidance on how you can increase your success in each of these areas.

Partnering with research and evaluation specialists

Throughout this handbook we emphasize that:

1. evidence should inform programme development;
2. evaluation should be incorporated into each step of programme implementation.

The former is critical to ensure that your programme is using best practice procedures, and by undertaking the latter – evaluation – you will be able to demonstrate which interventions work and which do not. By generating outcome data, you not only increase the likelihood of securing support from funding agencies but also help to secure the sustainability of your programme.

As previously mentioned, both these activities – research and evaluation – require specialist knowledge and skills. Specifically, researchers can help you with:

1. Identifying and summarizing high-quality research on risk and protective factors related to the child well-being concerns that your programme is addressing.
2. Identifying relevant evidence-based interventions, and helping you consider how they might be adapted for your communities.
3. Helping you translate your programme goals into measurable outcomes.
4. Assessing the pros and cons of different data collection strategies and evaluation study designs.
5. Identifying existing measures or indicators relevant to your project, and considering their relative merits.
6. Reviewing the ethical issues involved in measuring outcomes, and helping you ensure that children's and families' best interests and safety are incorporated into all stages of programme development and implementation.
7. Helping you plan for and execute your evaluation, including the collection of outcome data, data storage, data analysis, and the interpretation and reporting of results.
8. Registering your results if your evaluation is a randomized controlled trial in centralized registries so that others will be able to use and build on your programme's findings.

Collaborating with outside consultants or research partners

In practice, agencies or groups seeking to implement programmes to prevent children's exposure to violence vary in terms of the degree of research and evaluation expertise they have available in-house. Some agencies have no staff with research expertise, others may have one or two employees with research or evaluation backgrounds, while others have dedicated research and evaluation departments on their staff.

In the case of agencies and organizations that have research staff or departments, many of the activities listed above can be conducted in-house. For example, research staff can search

and summarize the scientific literature on an issue of interest, identify existing evidence-based practices, help programme development staff consider how evidence-based practices could be adapted to new contexts, and use research to inform programme theory models (see **SECTION 3: Defining your programme theory**). They can also be instrumental after an evaluation in using the results to make any necessary adaptations or changes to an existing programme or to develop a new intervention.

However, even if your agency has experienced researchers on staff, there are some very good reasons to seek outside partners to lead the evaluation process. There is now good evidence to suggest that positive bias occurs when outcome evaluation is conducted by programme developers themselves (Eisner, 2009). Even developmental or process evaluations may be influenced when conducted solely by internal staff, thus mitigating the benefit of the evaluation process. Although careful design, methodological rigour and open reporting of results can help to reduce bias, there will always be questions of impartiality when the developer, with a stake in the outcome, is conducting an evaluation of their own programme. One of the criticisms, for example, of the Triple P studies is that the majority of the evaluations of programme effectiveness were conducted by programme developers themselves (Wilson, Rush, Hussey, Puckering et al., 2012). Given you will be spending money on evaluation, it makes sense to connect with outside partners to ensure that nothing stands in the way of the acceptance of the validity of your results.

Aside from the issue of impartiality, there are a number of additional benefits to partnering with outside researchers or evaluators. Because of their distance, outside partners or evaluators can ask questions that those immersed in programme design may not have considered. They may feel less constrained in providing negative or critical feedback than those who are employed by your organization. Furthermore, such partnerships can add multi-disciplinary perspectives and new coalition-building opportunities. For instance, your partner may identify sources of funding for your programme that you were not aware of that you could apply for together.

Partnerships between programme developers and outside researchers are increasing in number and much is being learned about how to make such researcher–agency partnerships successful (Aniekwe, Hayman & Mdee, 2012). Arrangements that combine in-house expertise with the skills of an external evaluator – where, for example, agency researchers act as a liaison between the external researcher and programme implementers – appear to work well. Agency researchers can help make sure that the external research partner fully understands the programme, its objectives and the community in which it will be implemented; they can make sure that any differences in anticipated time frames, goals and perspectives are discussed and negotiated between the two partners; they can help facilitate data collection roadblocks or problems; they can convey the importance of the research to programme staff and participants; and they can help interpret the meaning of ambiguous findings.

An increasing number of foundations and funding agencies look favourably on applications which specify the involvement of external evaluators even in cases where in-house expertise is demonstrated; some now even require it. For example, a coalition of international foundations, who initiated the Children & Violence Evaluation Challenge Fund in 2011 to financially support evaluations of violence prevention and child protection initiatives, stipulate in their application process a researcher–agency partnership as a precondition to funding.⁷

⁷ For more info see: www.evaluationchallenge.org.

Finding an evaluation partner

There are a number of options for identifying external research and evaluation partners. You can contact potential funders and ask them for a recommendation. You could email researchers who specialize in the topic areas that are relevant to your programme; they might have students interested in helping to conduct an evaluation or know of colleagues who would be interested in your type of project. Alternatively, you could search for published papers reporting evaluations of similar programmes, contact the authors and ask them for advice.

If you are implementing a programme that has already been designed and evaluated in other settings, you may be able to interest the evaluators of that programme, their associates and/or their students in helping you set up your evaluation. The advantage here is that they would already have a detailed knowledge of the programme and probably an interest in seeing how the basic programme adapts to different communities or cultures.

Another option is to get to know the work of the academics at universities in the geographic areas in which your programme work will be conducted. Alternatively, you may be able to find a partner via one of the growing number of global evaluation associations and societies; such associations often have member lists or ways to connect you with an evaluator. Begin conversations about your interest in finding a research partner and explore the possibilities of working together.

Given that the skill level of researchers and evaluators does vary considerably, it is important to make sure that you get a sense of the type and level of research your potential partner has conducted in the past. The aim is to recruit someone with as much experience as possible in programme evaluation and in conducting research on issues affecting children and victims of violence. However, in some settings, especially in remote and under-resourced areas, it may be difficult to find local partners (e.g. a researcher at a local university) with this level of knowledge and expertise. One solution is a partnership between local evaluators and more experienced colleagues at other (more distant) universities who could provide oversight and consultation. Another option is to budget for training to increase the expertise of local evaluation partners in particular topics or evaluation methodologies.

Working with your evaluation partners

When embarking on a new working relationship with an evaluation partner you will need to decide what activities the collaboration will cover, how soon it will begin, and whether it will be limited to one project or continue across multiple projects. There is still much to be learned about what evaluation partnership configurations work best and what makes for a rewarding and productive working relationship. Nevertheless, it is unlikely that a successful evaluation can be conducted unless the evaluator is familiar with the work of your agency and its programmes, understands the implementation setting, the target population and its culture, and is willing to invest time in relationship building with programme staff. Consider talking with NGO staff or programme developers that your potential evaluation partner has worked with in the past to get an idea of their working style.

It is advisable to spend time early on with your own staff in order to identify possible areas of contention that could arise when bringing in outside evaluators, both in general and with the specific partner you have in mind. This will enable you to anticipate potential problems before they occur. Try to involve research partners in developing or refining your programme theory so that conflicting perspectives and priorities can be resolved at an early stage. In this

context, getting early agreement on the outcomes and outcome measures that will define the effectiveness of the programme is especially helpful.

Ideally, you will want to involve your research/evaluation partner in the funding application process, so that the required information about evaluation study design and outcomes is accurately and clearly specified. This handbook has been written in part to help facilitate that joint process.

As in any partnership, it is important to reach agreement in advance on matters such as the nature of the relationship, expectations about the goals and time frame of the work to be completed, and procedures for resolving problems. Be clear about deliverables – what outputs do you want from the researcher? Written reports? Published papers? Assistance with in-house capacity building? It is also vital to specify who will own the data after the project is completed; usually joint ownership is the most acceptable arrangement but discussing this issue directly early on will avoid later misconceptions.

Budgeting for evaluation

It is difficult to provide hard and fast guidance regarding the level of resources you should allocate to evaluating your programme. Some guidelines suggest setting aside a certain percentage of the overall budget to cover the costs of evaluation (e.g. 10–25%). However, the reality is that broad variations in evaluation study design and goals, in implementation costs, and in the types of evaluation partnerships make it hard to specify meaningful numbers. Using external partners is usually more expensive than relying on in-house expertise, and carrying out a rigorous evaluation will invariably cost more than a less rigorous one. However, funders, including many foundations, increasingly understand the value of the evaluation process and are accepting of the cost implications.

Ideally, your programme goals and your agency's interest in contributing unbiased evaluation data to the knowledge base should serve as a basis for meaningful back-and-forth budgetary discussions with potential funders. Discuss costs with your evaluation partner and use their expertise to identify alternative sources of funding that are less known to your agency and which could be used to supplement programme funding. Above all, make sure that you have sufficiently considered the full cost of the work. Think carefully and realistically about the costs of all of the key components of an evaluation, including planning, data collection, analysis and report writing. Some organizations have produced lists of potential budget items that may help you when drawing up a budget for carrying out an evaluation (W.K. Kellogg Foundation, 1998; see also **APPENDIX**: *Evaluation resources*).

Ethical issues

The field of child protection inevitably deals with sensitive, personal issues that involve stigma, shame and traumatic stress for many individuals who have been the victim of abuse in childhood. Furthermore, in some communities, disclosure of abuse is linked to a legal obligation to report that abuse to the appropriate authorities. These and other ethical considerations mean that when investigating violence and abuse outcomes, care and attention must be paid to participants' rights and needs at all stages of the evaluation process. In particular, you and your research and evaluation partners will have to put protections in place to ensure you collect data with the utmost sensitivity. Although this can be challenging, it should not be used as an excuse to avoid evaluation: learning what interventions work best to reduce abuse is the only way to secure a safer future for at-risk children.

All outcome plans and evaluation procedures will have to be reviewed very carefully, bearing in mind the ethical responsibilities to participants. The ethical issues involved in this kind of research are too complex to detail within the space of this handbook. Several resources are available that can guide your thinking through the ethical issues as you design your evaluation study (Child Protection Monitoring and Evaluation Reference Group, 2012; Finkelhor, Hamby, Turner & Walsh, in press).

You and your research partner will also need to identify what levels of approval will be needed for conducting this type of research, i.e. research involving children and their families and caregivers. The majority of universities, for instance, require ethics committee approval for any kind of research involving human subjects, and there are special requirements for research with children and when conducting research on sensitive topics. Make sure you understand how to get ethics approvals in your home country and, if applicable, in those countries where you plan to conduct your work.

Defining your programme theory

A programme theory is a specification of what changes you expect to observe as a result of your programme (i.e. your programme's outcomes) and how you think your programme interventions will lead to those changes. Defining a programme theory is a crucial first step in developing, implementing and evaluating a research-supported intervention. The process of creating and refining your programme theory will make it easier to choose appropriate outcomes and to design your evaluation. It will also help you explain to others, including potential funders, the logic behind what you want to do and how current research supports that logic.

Programme theory models go by many names, including “change models”, “programme rationale”, “logic models”, “pathway of change”, “theories of action”, “theories of change” or “causal theories”.⁸ There is a great deal of advice and writing on programme theory and logic models (see **APPENDIX: Evaluation resources**) using terms such as “interventions”, “inputs”, “determinants”, “outputs”, “proximal outcomes”, “distal outcomes”, “goals” and “objectives”, sometimes with slightly different meanings which can be confusing for those new to the concept. **BOX 3.1** provides a set of definitions of some of the more commonly-used terms in the programme theory model literature.

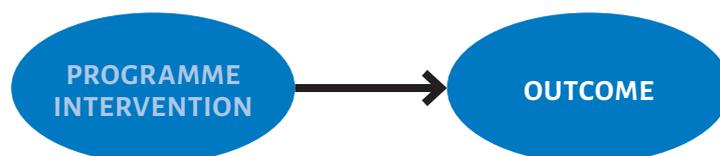
Programme theory: the basic concepts

Although some programme theory models are very complex, the most important elements are actually fairly simple. At its most basic, defining your programme theory means writing down:

1. the different components or elements of your programme, and
2. the specific and measurable outcomes you hope will happen as a result of those components.

One of the things that can help you better visualize the links in your programme theory is using a diagram or schematic to clarify how the elements connect to each other. A very simple schematic is shown in **FIGURE 3.1**.

FIGURE 3.1 A programme theory schematic



⁸ While these terms are often used to mean roughly the same thing, there are distinctions that are described as important by some authors (Chen, 2005; Funnell & Rogers, 2011).

BOX 3.1: DEFINITIONS OF “PROGRAMME THEORY” TERMS

Advice on constructing a programme theory model can be found in the many available books on programme evaluation (Chen, 2005; Funnell & Rogers, 2011; Patton, 1997; Rossi, Lipsey & Freeman, 2004; Wholey, Hatry & Newcomer, 2004). Additional resources are listed in the **APPENDIX: Evaluation resources**.

Below we provide brief definitions of some of the more commonly-used terms in the programme evaluation literature. However, keep in mind that those offering advice on how to define a programme theory often use these terms differently.

- **Inputs:** Financial, structural or staff resources needed or used by the programme to implement the interventions.
- **Activities:** Programme elements that will be implemented (e.g. interventions, treatments).
- **Determinants:** Also called causal, risk and protective factors, or mediating variables. These are deficits that the intervention hopes to reduce, or positive conditions and skills the intervention wants to build in order to achieve the outcome.
- **Outputs:** Usually defined in terms of services delivered or numbers of participants served; also called process or intervention data.
- **Outcomes:** Programme goals, objectives or benefits for participants or communities expected to result from the programme. Sometimes these are separated out into short-term, intermediate and longer-term outcomes. Outcomes are also sometimes also referred to as a programme’s **impact**. Some use the term “impact” to mean short-term objectives; some use it to mean longer-term or total programme effects.
- **Indicators:** These are the specific measures or data that are used to determine whether the programme outcomes have been achieved.

To illustrate the concept of programme theory models further, we will use a relatively simple example, based on a real-world evaluation of the effectiveness of children’s advocacy centres (CACs) conducted by Cross, Jones, Walsh, Simone et al. (2008). CACs, which currently operate in all 50 states in the United States, are agencies designed to improve a community’s response to child sexual abuse allegations. The key features of the CAC operational model are:

1. a multi-disciplinary investigation team (comprising, for example, representatives and professionals from child protection, mental health and medical services, NGOs, and criminal justice agencies and departments depending upon the community);
2. the use of specially-trained child forensic interviewers;
3. a “child-friendly” atmosphere for interviews;
4. procedures that link victims and families to mental health and medical services.

Note that the stated aim of a CAC is to “improve” the response to child sexual abuse. However, in terms of formulating a programme theory model, a programme’s outcomes (the right-hand side of Figure 3.1) need to be more precise than this. Thus, prior to planning the evaluation, staff at the participating CACs spent time reflecting on what specifically they, as programme developers, expected to look different as a result of their interventions.

Working in collaboration with an external evaluation team, agency staff determined that as a result of their programme they expected to see:

1. increased involvement of community agencies in child sexual abuse cases;
2. improved child and caregiver satisfaction with the investigation process;

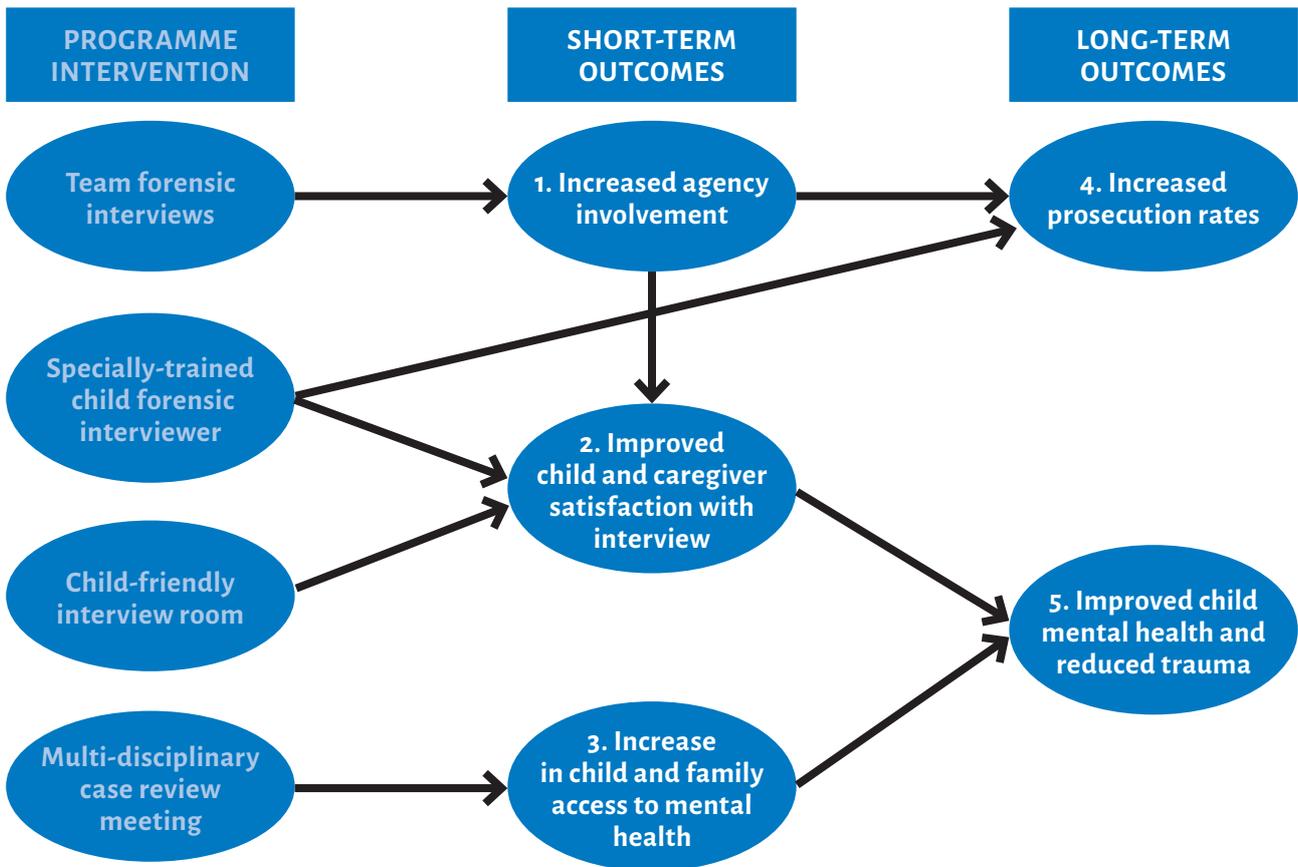
3. increased child protection rates;⁹
4. increased access of victims to medical and mental health services;
5. reduced re-victimization rates and improved mental health status among victims.

Having specified a set of programme outcomes, the next step was to consider the ways in which the CAC programme components (left-hand side of **FIGURE 3.1**) would influence the outcomes (right-hand side of **FIGURE 3.1**). These linkages were described by participating CAC agency staff as follows:

1. Team interviews would increase the number of agencies involved in child protection, and along with the child-friendly environment and specially-trained interviewers, would then improve child and family satisfaction with the interview process, and thus reduce traumatic stress for victims down the road.
2. Team interviews and the use of trained child forensic interviews would also, they hoped, result in improved protection outcomes (increased prosecution rates).
3. Multi-disciplinary meetings and specialized service agency agreements would increase children's access to needed services, and thus improve victims' mental health outcomes.

The resulting CAC programme theory model in schematic form is shown in **FIGURE 3.2**. In **BOX 3.2** we provide some additional tips to assist you in the task of formulating your own programme theory model.

FIGURE 3.2 Programme theory schematic for child advocacy centres (CACs)



⁹ In these settings, an increase in protection rates was more specifically defined as an increase in offender arrest and prosecution rates, but note that in communities without secure criminal justice protection system involvement, this outcome would need to be defined differently.

BOX 3.2: DEFINING YOUR PROGRAMME THEORY ELEMENTS: GENERAL TIPS

1. Define your programme components in concrete terms. For example: “Procedures to strengthen child protection systems” is too vague, as is “Provide protection to children whose rights are violated.” Precisely what procedures or protections are you going to implement? A parent education programme? A teacher education programme? A home visitation programme? Training programmes for law enforcement officers or doctors? A school-based prevention programme? Note that even these examples would need to be defined further. Being very specific about the interventions you are going to implement will help you clarify the outcomes you expect as a result, and will also help funding agencies understand what it is you want to do.
2. Likewise, your outcomes need to be as specific as possible. It is usually a good idea to include short-term, intermediate and long-term outcomes as you begin the process of defining your programme theory. Make sure that you are clear what programme element is expected to affect which outcome.
3. You can make the programme theory model as complicated as is helpful to you during brainstorming sessions, but bear in mind that the goal is to define the intervention–outcome pathways that you think are the most important, so keep it **functional, realistic and measurable**. In particular, when drafting a programme theory model for funding purposes, make sure that the outcomes you include in your model are measurable within the limits of project funds and time frame. Choosing appropriate outcomes is discussed in more detail in **SECTION 4** of this handbook.

Extending your programme theory model to include causal and risk factors

While the basic model framework outlined above is a good place to start, a programme theory model improves further by including statements about *why* you think your intervention will result in your specified outcomes. In other words, it is helpful to think about the underlying conditions or behaviours – the risk factors – that you think are causing the problem that you are seeking to address with your programme interventions and include these in your model; see **FIGURE 3.3**.

FIGURE 3.3 Including causal factors in your programme theory schematic

The main advantage of including causal factors in your model is that it will help you use existing research literature to check the assumptions that underpin your programme theory. Several cautionary tales – in which interventions were developed and widely implemented without first checking that programme theory assumptions were supported by research findings – serve to highlight the importance of this step. Early child sexual abuse prevention efforts in the United States, for example, were based on a faulty understanding of the problem. It was assumed that child sexual abuse was primarily an issue of stranger abduction and this assumption led to misguided education efforts that were later found to be ineffective. Better results were achieved once the dynamics of the main problem were understood – that most children are molested by people known to them (Finkelhor, 1979).

Similarly, early drug abuse prevention programmes presumed that young people did not understand the risks associated with drug taking. The underlying programme logic supposed that if youth were educated about those dangers, they would avoid drugs. However,

subsequent studies revealed that educational approaches based on this “information deficit” model had little effect. It wasn't until educators recognized the importance of peer influence, and developed interventions to help youth resist negative influences, that it was possible to design more effective prevention programmes (Lantz, Jacobson, Warner, Wasserman et al., 2000).

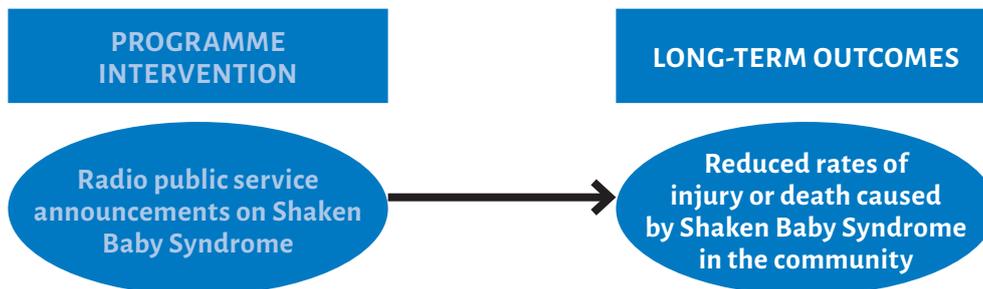
Note that in Figure 3.3, we placed the causal factor in the middle of our schematic to illustrate the idea that a programme intervention is designed to *influence or change the causal factor* and thus bring about a change in an outcome. Some models place the causal or risk factors first – to the left of the programme components – to signify the environment in which the programme intervention occurs. It doesn't really matter how you organize your programme theory model, as long as the assumptions and pathways are clear.

Below we use two hypothetical examples to illustrate the ideas behind the use of an extended programme theory model further.

Example 1: A public service announcement to reduce Shaken Baby Syndrome

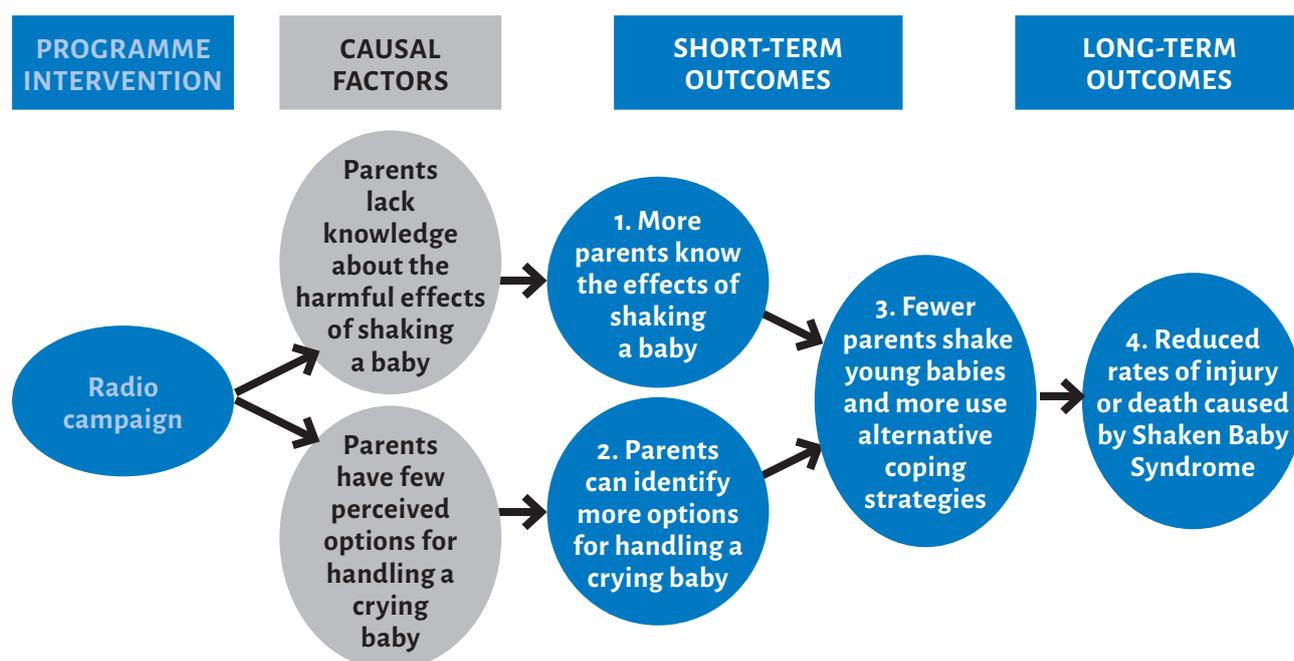
Consider a situation whereby programme developers wish to set up a public awareness campaign in several communities to reduce the incidence of Shaken Baby Syndrome (SBS). They draw up an initial programme theory diagram: see **FIGURE 3.4**.

FIGURE 3.4 Initial programme theory schematic for a hypothetical programme to reduce Shaken Baby Syndrome



The programme developers wish to extend their model by adding causal factors. They decide that their intervention is based on a belief that two specific problems exist: 1) young parents in the target communities do not understand that shaking a baby can result in traumatic injury or death; and 2) young parents are not able to identify alternative, safer options for handling their anger when their baby cries. The programme intervention – the public service announcement – now has two specific goals: 1) educating parents about the physical effect of shaking babies, and 2) providing them with some alternative strategies when they are angry or cannot get their baby to stop crying. **FIGURE 3.5** shows the resulting, more developed, programme theory model for the media campaign.

It is important to make sure that there is evidence available to support the causal links you are hypothesizing in your model. If programme developers in our hypothetical example can cite research that parents' lack of information and awareness is related to the problem of shaking infants, they can feel comfortable that their programme theory model is sound. Suppose, however, their research review indicated that uncontrolled anger is a more important causal factor than a lack of information. In this scenario, a more effective approach might be to

FIGURE 3.5 Hypothetical programme theory schematic including causal factors

offer home-based anger management education to young or at-risk new parents rather than to rely on a radio campaign.

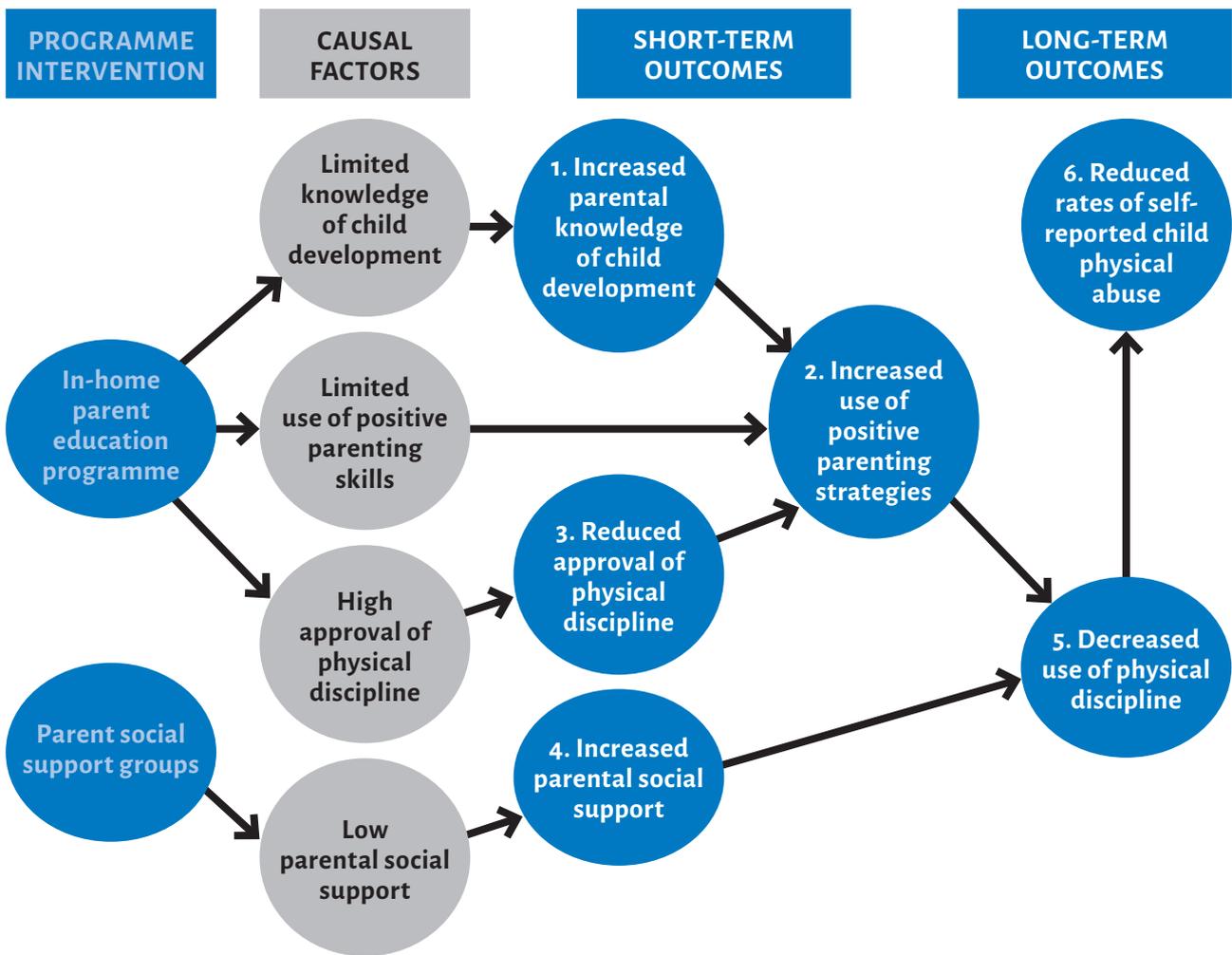
In addition to confirming whether your causal assumptions are valid, reviewing background research while developing your programme theory might also lead to the discovery of existing evidence-based interventions targeting the same concern. For instance, a literature review of research on Shaken Baby Syndrome would likely identify the Period of Purple Crying Programme (Barr, Barr, Fujiwara, Conway et al., 2009; Barr, Rivara, Barr, Cummings et al., 2009). This evidence-based educational programme, designed to be implemented in maternity settings and through community-wide media campaigns, informs parents about normal infant crying and the dangers of shaking a baby. In learning more about this programme, the programme developers in our example would have the option of implementing the Purple Crying programme “as is” or adapting elements of it, depending on how well they think it would translate to the communities they are working with.

Example 2: A parent education programme

As a second example, we consider a hypothetical parent education programme aimed at reducing child physical abuse rates. Suppose that in this case, the programme developers have determined, with assistance from their research partners, that child physical abuse is greater when parents’ knowledge of child development is low, their knowledge of positive parenting strategies is limited, they approve of physical discipline, and levels of social support are low. They decide to target these risk factors with a community programme that includes: 1) an in-home parent education intervention, and 2) parent support group meetings held in the community.

The schematic of the resulting programme theory model (FIGURE 3.6) shows that the programme developers hope to achieve higher levels of child development knowledge among parents, increased use of positive parenting skills, reduced approval of physical

FIGURE 3.6 Programme theory schematic for a hypothetical parent education programme



discipline and increased social support for parents (outcomes 1–4). They hope this will lead to decreased use of physical punishments (outcome 5) and reduced levels of child physical abuse (outcome 6).

This example provides a simplified version of what interventions a parent-based education programme might include in practice. For instance, having determined from research that mental health and substance abuse issues are significant risk factors for child physical abuse and neglect, programme developers might opt to provide services to help parents with these problems. Others might focus on interventions to improve parent–child interactions. These sorts of decisions will determine the programme outcomes, and depending on the focus, lead the programme teams to different sets of short-term, intermediate and long-term outcomes.

Regardless of the specifics of the programme interventions, the overriding goal is to make as explicit as possible the linkages between the intervention strategies, the causal problems being targeted, and the expected outcomes. The next step is then to assess how well the programme theory is supported by existing research. This is latter task discussed below, and also in **BOX 3.3**, which provides additional tips on using research to improve your programme theory.

Finding the research on your programme theory elements

As a programme developer, it is important to have a good understanding of the current research in the field in which you are working. Several resources are available to help you achieve this aim. Summaries of research on different issues related to child violence and protection are widely available as published review articles or book chapters. An Internet search can provide a useful synopsis of the research literature on a given topic. There are also many websites that showcase the work of research centres and researchers who study children's exposure to violence. You can visit these and find out if they have conducted research on the particular issue you are interested in (see **APPENDIX: Evaluation resources**).

However, as was discussed in **SECTION 2: Partnering with research and evaluation specialists**, it is advantageous to use experienced researchers to help you conduct more targeted reviews of particular research questions. The reasons for this are:

1. The quality of research studies and articles varies a great deal. Someone with a background in research and evaluation can identify those studies that have a stronger methodology. They will be better able to determine which findings have the most significance for what you want to accomplish with your programme.
2. For some less well-studied topics, published research may be difficult to find. Conversely, for well-studied areas, the availability of a large amount of research material can be overwhelming. An experienced researcher will have the skill set to uncover hard-to-find sources of data or, when research on a topic is plentiful, narrow the search down to the most critical questions.
3. Finally, conducting a good literature review can be a time-consuming exercise. A research consultant will be able to carry out such a task more efficiently than you, summarize the key points and, most usefully, discuss with you the implications for your own programme goals.

BOX 3.3: TIPS FOR USING RESEARCH TO DESIGN AND SUPPORT YOUR PROGRAMME THEORY

- It is important to be honest about what you find. If research shows that an anticipated programme theory link is false (e.g. using fear-based tactics does not reduce peer-violence), then you will need to adapt your programme and develop interventions that fit better with the research.
- Make sure you review evaluations of interventions that have already been tried in your area of work, especially those which have reported positive results. It will be much easier to argue that your programme is research-supported if you build on an intervention with proven efficacy. If you are not sure the proven intervention will translate well to the communities you want to work with, then develop adaptations. Certainly adapting interventions to new contexts, especially between countries with different cultural norms and levels of socioeconomic development, is not an easy process. There is much still to be learned about the circumstances in which such adaptations work and when they do not (Barrera & Castro, 2006; Castro, Barrera Jr & Holleran Steiker, 2010; Cuijpers, De Graaf & Bohlmeijer, 2005). Nonetheless, there is much that does translate across cultures, and starting with what has worked in one community can speed up the work in finding the right intervention for another.
- Sometimes there will be no available research on some, or all, of the hypothesized links in your programme theory model. In this situation you will need to acknowledge that your approach is innovative and untested, and evaluation is particularly critical in these circumstances. However, you will still need research to back up your argument that your innovation is needed and likely to work.

Choosing and measuring outcomes

Once you have researched and defined your programme theory, you will be well on your way to creating a research-supported programme and you will have established a good basis for arguing its likely effectiveness. However, it is still necessary to evaluate your programme to check that it is making the difference you think it should. By so doing, you will be adding to the evidence-base and giving others the opportunity to learn from and build on your work.

The first step in the evaluation process is defining programme outcomes. Well-defined programme outcomes share two main characteristics:

1. they are specific, i.e. they describe the specific changes you want to see in the circumstances of the children and families that you are trying to help;
2. they are measurable.

As we noted in Section 3, terms such as “outputs”, “impacts”, “short-term outcomes”, “long-term outcomes”, “indicators” and “measures” are used by different evaluation experts to mean slightly different things. The terms matter less than being clear about their meaning. For the purposes of this handbook, we define these terms as follows:

- **Outputs.** In our handbook, “outputs” refers to the information that tells you if your programme has been implemented successfully as planned. This would include information like the number of meetings held, clients served, people who attended your sessions or professionals trained.
- **Outcomes.** By “outcomes”, we mean the effects of the programme on children and their families, for example, a reduction in physical abuse or neglect rates. As we demonstrated in the previous section, distinguishing short-term and long-term outcomes can help you to better understand your programme theory and plan your evaluation in terms of which outcomes you can feasibly measure.
- **Indicators.** We use the term “indicators” to refer to the specific information (e.g. scale or index scores, administrative data) that you can use to track outcomes.

Defining outcomes: common pitfalls

Defining workable programme outcomes and selecting appropriate indicators is not an easy process. Common mistakes that some programme developers make are described below.

Positive endorsements of the programme

As we mentioned in the first section of the handbook, positive feelings about the programme reported by recipients, anecdotes about how the programme has helped children and families, and professional opinions about the programme do not provide a basis for convincing outcome data. Collecting information on recipients’ experiences might

be valuable for informing programme development, and if done well, can be a critical part of developmental or process evaluations (see **SECTION 5** for more details). Client feedback data are useful in understanding the nature of service delivery problems and can help you adjust certain programme elements to ensure that they run smoothly.

However, as a rule, positive endorsements – even when collected using a structured survey – do not constitute good “outcome” data. This type of information does not do a reliable job of telling you whether or not your programme has been effective in achieving its goals. Any programme, good or bad, can find these kinds of endorsements. Clients or programme recipients, particularly those who have few resources or little support, tend to have a low threshold for what counts as “helpful” which can lead to a positive skew in the results of an evaluation of programme performance based on participant satisfaction data (Stallard, 1996). Also, just because a programme has been perceived positively does not necessarily mean it has led to a reduction in children’s risk for violence or an improvement in protection services.

Listing implementation goals as outcomes

Avoid stating your programme’s implementation goals as outcomes. Implementation objectives such as “the establishment of a programme to protect children”, “a network of organizations created to help protect children” or “stronger child protection systems” may well be valid outputs in your programme theory model and may even form end objectives for a process evaluation (see **SECTION 5**), but they are not useful as programme outcomes. They do not provide information on how well your interventions are doing in terms of improving the safety and well-being of children and families.

To reinforce this point, consider these examples:

- ✚ Having improved professional networks does not necessarily mean that more children are protected (because we don’t know whether the professionals took steps to increase child safety).
- ✚ Providing more mothers with parent training does not necessarily equate to a reduction in abusive behaviours (because we don’t know how good or appropriate that training was).
- ✚ Training more mental health providers does not demonstrate that more children have access to mental health services or that children’s traumatic stress levels have been reduced (because we do not know how many children visited mental health services or whether they felt better as a result).

In each case, outcome measurement is needed to show that the end successes have actually been achieved.

Using vague or ill-defined outcomes

Another mistake that programme developers sometimes make is defining outcomes that are too “big” or too vague. **TABLE 4.1** lists examples of outcomes that are only loosely defined (left-hand column) and translates these into more specific outcomes (right-hand column). Notice that the outcomes on the right lend themselves much more easily to measurement.

The outcomes in the left-hand column of **TABLE 4.1** may be good starting points for discussions about your programme theory. Most programmes hope for powerful and significant long-term impacts, such as improved safety and well-being among children in the community, but it is likely to be more helpful if you think of these as overarching programme

TABLE 4.1 Examples of vague versus well-defined outcomes

OUTCOMES THAT HAVE NOT BEEN WELL DEFINED	WELL-DEFINED OUTCOMES
Improved child well-being in the community	Decreased levels of post-traumatic stress in children one year after programme delivery
The community is mobilized to protect children	Increased percentage of family or community members describe making a report to a designated protective organization
The community is better educated about child abuse	Decreased endorsement of violence against children as an acceptable disciplinary strategy
Reduced victimization of children in the community	Decreased rates of physical punishment by teachers in schools after receiving the programme
Domestic violence is no longer a problem in the community	Decreased rates of domestic violence victimization reported by programme recipient partners within 6-months of programme delivery compared with those who did not receive the programme

goals. Your research partner(s) will be able to help you translate your programme goals into measurable outcomes, and to present these to potential funders.

Outcomes that are unrealistic in terms of the time frame of the programme

It is also important to select programme outcomes that you anticipate will measurably change within the time frame of your evaluation. Consider what changes you expect to see as a result of your programme at various time points, such as immediately after programme delivery, 1 month later, 6 months later, 1–2 years later and after 5 years. How do you hope the problems you are addressing will look at each of these points?

An ambitious programme may plan to improve maternal financial independence, providing opportunities for their children to advance economically, thereby decreasing the later risks of violence associated with poverty, but this is going to be difficult to demonstrate within most evaluation time frames. Selecting outcomes means not only considering what you feel is important about a programme's impact, but also what can be measured within a given amount of time.

Outcome measurement strategies and tools

The term “indicator” is typically used to describe the outcome measure or “yardstick” that you will use to determine whether or not your stated outcomes have been achieved (see above definitions). Indicators can be derived from questionnaires, scales or from data routinely collected by communities, agencies or governments and stored in registries or official records. For instance, if your outcome is “improved child mental health”, your selected indicator might be depression scores as measured by the Child Behaviour Checklist (CBCL). If your outcome is “decreased physical punishment by teachers in school”, your corresponding indicator might be rates of physical discipline used by teachers in the previous month, either self-reported by teachers in a survey questionnaire or based on classroom observations made by researchers.

Choosing appropriate outcome measurement tools or indicators is another area where the input of a partner with a research and evaluation background will be an advantage. Below we briefly review some the options for indicator selection that you can discuss with your evaluation partner or consultant when planning your programme evaluation.

Survey data

A benefit of using survey instruments (e.g. questionnaires) to collect outcome data is that you are able to choose what you want to measure and you have control over *how* it is measured. You are also getting information directly from your programme's recipients, the children and their families themselves. Survey instruments are very versatile: you can use them to measure mental health or trauma symptoms, to assess parenting skills, to test knowledge, or to ask people about their victimization experiences. If you are planning to use survey data you will need to consider whether you will rely on existing assessment measures and tools (e.g. scores, indices, scales) or construct your own, and decide who you want to take part in the survey (i.e. your sample) and at what time points.

Creating good assessment measures from scratch is difficult. It is always worth finding out if there are any existing scales or indices that would be appropriate for the outcomes you are trying to measure. You should be looking for measures that have strong psychometric support (i.e. are standardized, and with evidence of reliability and validity). The US-based Centers for Disease Control¹⁰ has compiled several compendiums that provide comprehensive lists of standardized assessment measures on topics such as attitudes about violence and victimization experiences.

Alternatively you can go back to the research review you completed when developing your programme theory and see if the literature mentions any measurement tools that you could use or adapt. Some structured scales and indices cost money, but many are freely available in the public domain. Many researchers are open to providing others with the measures that they have developed and used in their research. There is an increasing interest in translating and testing well-established assessment tools in countries around the world, and your efforts can contribute to this goal.

If you do need to create your own measurement tool, or adapt an existing one extensively, we strongly advise you to work closely with your researchers or research partner on this task. Scale development is a very complex process and it is easy to make design mistakes that can result in uninformative outcome data.

Observational data collection methods

In situations where collection of information directly from individuals is not feasible or researchers prefer not to rely on self-reported data, observational data may be an option. For example, trained research staff might record youth and school staff behaviours in schools, or observe and document parent–child interactions in the home or at a clinic. Observational data can be a valuable source of information on families and children, although it is labour and resource intensive to collect.¹¹

It is important for programme developers to be familiar with the strategies that ensure that observational data are collected in valid and generalizable ways. One such strategy is blinding. A good observational study design will involve the use of “blind” data recorders – observers who do not know whether the subject is receiving the intervention or not. Blinding is an accepted way of minimizing observer bias, ensuring that observers do not bias their observations based on their expectation of a treatment effect.

¹⁰ See for example: http://www.cdc.gov/ncipc/pub-res/pdf/YV/YV_Compndium.pdf

¹¹ For more information on the use of observational data, go to: <http://www.cdc.gov/healthyyouth/evaluation/pdf/brief16.pdf>.

Administrative data or record review

Some of the outcomes that you might be interested in can be measured by collecting administrative data from clinical, agency or governmental records, or from surveys of professionals. For example, it may be possible to measure an outcome like physical abuse victimization by collecting information on abuse-related injuries from local hospitals.

Administrative data are usually relatively easy to obtain in communities where such data are routinely collected and kept in a central location. However, in many settings, especially in low- and middle-income countries, administrative data are either not available at all or are unreliable. Across all communities, administrative data can be limited in coverage, badly organized or missing critical information, making it difficult and time consuming to extract the relevant information from them. For these reasons, use of agency records as a source of outcome data requires specialist analytical techniques (e.g. for handling missing data) and the help of a suitably skilled research partner.

If you are planning to use agency-based indicators in your evaluation, it may be worth building relationships with the key agencies involved in data gathering so that together you can improve the quality of the data. Better record-keeping on all issues related to child health and safety is an important goal for any community, given that enhanced information systems facilitate service tracking and continuum of care models for child victims or at-risk youth.

Qualitative data

Qualitative data, such as that generated through interviews and focus groups, are a fourth option for collecting information on outcomes. In general, qualitative information is less objective, relies on smaller samples and is harder to interpret than quantitative data. Nevertheless, qualitative data can be instructive, especially when used in conjunction with information on other measurement tools. For example, you could use qualitative data as a supplement to quantitative measures in a mixed-method evaluation study design to provide deeper understanding about how and why changes did or did not happen as a result of the implementation of your programme.

Additional considerations

Using proxy indicators

For programme developers and implementers seeking to reduce the prevalence of child abuse, child neglect, sexual assault and exploitation, and other forms of children's exposure to violence, one of the most important outcomes will be a reduction in violence exposure for children in the target population. However, measuring reductions in abuse, aggression or violence, particularly as it is frequently perpetrated within families or by caregivers, is difficult. There are ethical issues to consider when asking individuals to self-report abusive behaviour or victimization and the sensitivity of the issue demands careful survey and measurement approaches. In addition, some types of violence and aggression occur with relatively low frequency, making it difficult to monitor change. There are also data quality problems associated with the use of child protection or criminal justice administrative records as a source of prevalence data on violence and abuse as we describe above.

Despite these measurement challenges, much good work has been done and continues to be done to create survey methodology and tools to measure levels of victimization. Similarly, even with their limitations, official government child protection or child abuse arrest reports remain valuable sources of information and work is ongoing to improve their availability and

quality in many countries. These efforts are critical to identifying what kinds of programmes reduce children's exposure to violence and abuse.

Nonetheless, given the difficulties in measuring and detecting a difference in violence exposure outcomes, researchers often include proxy indicators in an evaluation study. These are indicators that, according to research, are very closely associated with rates of abuse – either as correlates, causes or direct consequences of violence exposure. Depending on the type of violence exposure being targeted, examples of suitable proxy indicators might include: rates of parental physical punishment or harsh parenting; parenting stress; the use of positive parenting strategies by parents; school attendance record; child health and mental health status; and achievement of child developmental milestones.

Need for comparison data

Although we discuss evaluation methodology in more detail in the next section, we make the observation here that in order to understand whether changes in your outcome indicators can be attributed to your programme and your programme alone, you will need to collect some kind of comparison data. You can do this by collecting data pre- and post-intervention, collecting data from comparative samples (i.e. children, families or communities not receiving your programme – your control group), or both.

Some programmes collect outcome information by asking participants to complete a questionnaire at the end of a programme, and then checking to see how many were able to answer follow-up questions “correctly” in order to determine how much participants had learned. However, without a point of reference or comparison, the proportion of correct responses achieved post-intervention does not tell you that much about programme performance. If participants score poorly, and fail to get a reasonable proportion of right answers, then that is certainly a bad sign. But right answers could mean a number of things: the questions may be obvious or easy to guess, recipients may have known the right answers before the programme, or they may have obtained the necessary information from some other source. All of these scenarios would suggest that your programme is performing well, when in truth it might not be.

Sample selection

Depending on the size of the target population, outcome data do not necessarily have to be collected from everyone who participates in your programme. You will, however, need to use careful strategies to choose who outcome data will be collected from. By selecting your sample as randomly as possible, you reduce the potential for selection bias. Selection bias arises, for example, when a survey is administered to those that are easiest to reach or that seem to have benefitted the most from the programme.

In addition to sample selection procedures, sample size will also be an important issue to consider with your research partners. Generally speaking, the larger the sample you have, the more confident you can be that your outcome data will represent the experiences of typical children and families. Having an adequate sample size also ensures that you have enough research “power” to detect the changes you think will happen in your population of interest as a result of your programme. Ultimately, the optimal sample size will depend on the nature of your intervention, the numbers of people involved in the programme, and your evaluation study design. Your evaluation partner can help you make sure that you are planning to collect outcome data from sufficient numbers of subjects and using appropriate sampling strategies.

Understanding programme evaluation

In the sections 3 and 4 of the handbook, we discussed developing a programme theory model and creating a schematic to communicate this model to others, as well as options for defining and measuring programme outcomes. These activities are essential preparation for planning and conducting an evaluation of your programme. In this final section we provide a brief summary of different types of evaluation and their study designs, address some common questions, and highlight some important issues to consider when planning your evaluation study.

What does it mean to have an *evidence-based* programme?

Evidence-based programmes and interventions are something that funding agencies and governments are increasingly seeking. But what does it mean to be “evidence-based”?

The term evidence-based is sometimes used to describe a programme that has been developed using existing research to inform its design. We suggest that it is more helpful to designate such programmes as “evidence-supported” and reserve the term “evidence-based” for programmes that have some direct evidence of their effectiveness. In other words, an evidence-based programme is one that has gone through *a rigorous evaluation process which has demonstrated effectiveness for at least some of the key intended outcomes*. In **BOX 5.1** we list six elements that define a rigorous evaluation. The more of these six elements your evaluation contains, the more certain you can feel that your results are real and meaningful.

BOX 5.1: SIX KEY ELEMENTS OF RIGOROUS EVALUATION RESEARCH

The most rigorous evaluations include the following six key elements:

1. Recipients of the programme (the “treatment” group) are compared with a group of individuals who do not receive the programme (the “control” group).
2. Recipients (or communities) are randomly assigned to either the treatment or control group.
3. The study sample is both sufficiently large and representative of the target population to allow generalization of the results to the population of interest with a reasonable degree of confidence.
4. Programme delivery is administered and monitored in a way that evaluators can be certain that it was implemented as intended.
5. The outcomes are not only studied immediately after the intervention is administered but also at later points in time, to make sure that any effects are sustained.
6. Outcomes are not limited to changes in knowledge and attitude, but include changes in behaviour and experiences, such as decreased victimization rates.

It is an intensive endeavour to conduct an evaluation that includes all six of the elements listed in **BOX 5.1**. Such studies require significant resources and the involvement of highly skilled research experts. Evaluations falling short of this ideal can still produce valid, meaningful information. For example, it might not be possible to randomize assignment, but other techniques that can be used to try to make the treatment and comparison group as equivalent as possible. Furthermore, if longer-term follow up is not possible, the results can still be interpreted as indicating that the programme is effective “initially”. Finally, while behaviour change is often the ultimate goal, evidence of sustained change in knowledge and attitudes can be an acceptable sign of programme impact as long as the limitations of this approach are made clear.

While including all of the six elements in **BOX 5.1** is not required for an evaluation to be useful, it is important to appreciate how and why the absence of any one of the elements limits your understanding of your programme’s impact and reduces the strength of the evidence for the effectiveness of your programme. Although it is beyond the scope of this handbook to discuss the significance of each of these six elements in detail, programme developers and implementers should have at least a basic knowledge of why things like control groups, random assignment and adequate sample size are important. The resources listed in the Appendix can help you learn more about these matters. We recommend discussing each of these important methodological aspects with your evaluation partners and incorporating as many as you can.

Preparing for evaluation

Given that evaluation is both resource- and time-intensive, you will want to make sure – as far as is practically possible – that your evaluation efforts do not fail because of poor preparation and lack of “readiness”. The tasks we described in Sections 3 and 4 – developing a sound programme theory and specifying measurable outcomes – form part of the work you will need to do to prepare for evaluation. Some researchers refer to a formal process for checking for evaluation readiness as an “evaluability assessment” (see, for example, Justice Research and Statistics Association, 2003). Such a process can help you and your research partner determine whether a programme (and the community in which it is to be implemented) is ready for an evaluation of its effectiveness.

Developmental and process evaluations (described below) can provide information on a programme’s readiness for evaluation by helping to ensure that the programme is being implemented as intended and delivered to the expected recipients. Evaluability assessments often also confirm that the agency or setting in which the programme is being delivered is prepared to assist with the necessary data collection procedures. Such checks enhance the likelihood that outcome data for the programme will be successfully collected.

Different kinds of evaluation

Different kinds of evaluations are useful at different stages of programme development and implementation, and each provide different types of information. Below we distinguish two main types of evaluations: process evaluations and outcome evaluations.

Process evaluations

The primary aim of process evaluations is to collect information which can be fed back into programme development and refinement. “Developmental evaluations”, which are similar, are evaluations that happen very early on in a programme’s development cycle,

whereas process evaluations can take place at any stage of programme development and implementation.

Data generated by developmental and process evaluations can be extremely varied and can include:

1. implementation data (information about how the programme is being implemented and which individuals or groups are receiving which services);
2. interview or focus group data, or survey data from clients, staff, partner agency staff and the community (information about how programme implementation is going and how the programme is being used and perceived by stakeholders);
3. outcome data (information about changes in risk and behaviour for the children and families receiving the programme).

All of this information can be used to modify and refine the programme and its delivery. Because you are not trying to make a statement about the impact of the programme on recipients compared with non-recipients, the methodological rigour needed for outcome evaluations is less stringent. A good example of a well-conducted and well-reported process evaluation of a family-strengthening intervention in Rwanda is provided by Betancourt Meyers-Ohki, Stevenson, Ingabire et al. (2011).

Impact or outcome evaluations

In contrast to process evaluations, outcome evaluations provide you with information on whether your programme is achieving the expected outcomes for the population being targeted. This kind of evaluation requires stricter rigour than process evaluations, because the purpose is to show that receiving the programme is better for children and families across key outcomes than not receiving it. The methodological structures inherent in these kinds of evaluations are the best guarantees we have against erroneous conclusions and the tendency to see patterns and positive effects where they don't exist.

Types of outcome evaluation

There is a wide range in sophistication, size and rigour across outcome evaluations. While all of the study designs reviewed below can provide information on programme effectiveness, keep in mind that including the elements cited in **BOX 5.1** (*Six key elements of rigorous evaluation research*) increases confidence in the findings.

Pre-post comparison designs

We previously explained that, because of the lack of comparison information, post-intervention-only data collection will not tell you very much about the impact of your programme (see *Need for comparison data* in **SECTION 4**). However, by comparing outcomes prior to and after the introduction of an intervention you add some rigour to the post-test only design. For instance, you might do a survey of a community or neighbourhood before and after implementing an educational or public awareness programme. Alternatively, you might collect data on families, children or other clients before and after you implement a new service.

Pre-post comparison study designs can tell you how programme recipients' knowledge, behaviour and/or experiences differ after receiving your programme intervention. Your understanding of the sustainability of your programme's impacts will be enhanced if you are

able to repeat data collection at several intervals post intervention. However, it is important to be aware that there may be many reasons why outcomes change over time for programme recipients that have nothing to do with your programme (e.g. parenting skills may increase or victimization rates may fluctuate naturally over time). In order to be able to attribute the measured changes to the programme interventions with any degree of confidence, comparison information from a control group is needed.

Quasi-experimental designs

In a quasi-experimental study design, outcomes in the treatment group—the people receiving the programme—are contrasted with those of a control group—a group of people who are as similar as possible to the treatment group but who are not exposed to the programme. Any observed differences in outcomes can then, in theory, be attributed to the effect of the programme.

Depending on the programme or intervention, the selection of a suitable comparison group can be problematic and often creative solutions are needed. It might be possible, for example, to use people who are on a waiting list for receiving the programme or intervention being evaluated as a control group, particularly if the programme accepts people on a first-come-first-serve basis. Alternatively, evaluators can try to identify a similar neighbourhood, school or community that will not be receiving the programme and recruit a comparison group from that location. Often communities and schools will let you collect data if they think it might help them gain similar services in the future should the programme prove to be successful.

The biggest problem with quasi-experimental study designs, and the reason that randomized controlled trials (described below) are considered preferable, is that without formal randomisation procedures, it is very difficult to get two groups that are comparable in every way except for the experience of an intervention. Schools, neighbourhoods and communities are likely to differ by problem incidence or severity; they may also differ materially in terms of demographics, socioeconomic status and resource availability. All of these factors can affect outcomes in ways that make it hard to be sure whether a observed change can be attributed to the effect of your programme, or to some other group difference.

The quasi-experimental design can be improved if you are able to also conduct pre- and post- surveys on both groups and compare outcomes before and after the introduction of the intervention; this will allow you, through the application of various statistical techniques, to account for some of the pre-existing differences in outcomes between the two groups. Regardless, it will be necessary to make a case for the similarity of the groups as best you can, while recognizing the limitations of this approach relative to a more rigorous RCT-based study design.

Randomized controlled trials

The study design that provides the highest level of rigour in evaluation is the randomized controlled trial (RCT). With this design, questions of differences in the treatment and control groups are significantly reduced by randomly assigning individuals to the group that receives the programme and the group that does not. In the case of interventions that are difficult to administer on an individual basis such as a school-based peer violence prevention programme, you can randomly assign whole classrooms, schools or households to either the treatment or control group. This study design is known as a cluster randomized controlled trial.

Many programme developers are uneasy with the idea of withholding – at random – a programme or a service that they anticipate will be beneficial. They may feel ethically obligated to either give the intervention to everyone, or to those most in need. However, it is important to keep in mind that without outcome data, it cannot be known for sure whether, or to what degree, the programme really benefits recipients. In some cases it may be possible to mitigate the sense that random assignment is unfair by offering those in the control group alternative, standard or “status-quo” services for example, and/or provide assurances that they will also receive the new programme or services as soon as the study period is over.

Some researchers have expressed their concern that there is too much emphasis placed on RCTs as the “gold-standard” in evaluation. It is true that some programme delivery models or services do not lend themselves to evaluation using a RCT. For example, very complex system-level or policy change efforts are not easy to “randomly assign” to different communities. It is also impractical to randomly assign programmes that require multi-agency involvement and significant resources to implement. In some cases, a very well conducted quasi-experimental study might be the only viable option. Nevertheless, the RCT is a gold-standard for good reason – it is the design that provides the greatest protection against erroneous conclusions. If you wish to demonstrate with the utmost confidence that your programme has been effective, you will need the rigour provided by a randomized controlled trial.

Despite the challenges, an increasing number of RCTs are being successfully conducted, and producing informative results, even in less well resourced, low- and middle income countries (see **BOX 5.2**). The value of RCTs to our knowledge of effective prevention and intervention strategies can be seen in the growing number of clinical trial databases. These registries ensure that the results of all studies (even unpublished ones) can be reviewed and used by researchers. A list of trial databases of relevance to child protection are provided in the **APPENDIX: Evaluation resources**.

BOX 5.2: AN EXAMPLE OF A CLUSTER RANDOMIZED CONTROLLED TRIAL IN JAMAICA

A cluster randomized controlled trial of the Incredible Years Teacher Training intervention was conducted in Kingston, Jamaica in order to evaluate its impact on the behaviour of high-risk children (Baker-Henningham, Scott, Jones & Walker, 2012). Given the generally positive evaluation findings in several high-income countries, the authors were interested in seeing if similar results could be achieved in low- and middle-income countries. The programme, which trains teachers in classroom management skills (such as building positive relationships with students, using positive discipline techniques, motivating children with incentives, and teaching children emotional regulation skills and problem solving) was tailored by the researchers for delivery in Jamaican pre-schools (Baker Henningham, 2011) and then piloted (Baker Henningham, Walker, Powell & Gardner, 2009).

The trial itself involved randomly assigning 24 pre-schools in Kingston to either a treatment or control condition. Data on children's behaviour were collected via observation; parent and teacher report data were also collected. The results of the study showed that children in the intervention schools had fewer conduct problems and increased friendship skills compared with their peers in non-intervention schools.

This study provides a good example of the approach advocated in this handbook – using and building on existing research to improve the evidence-base. The researchers selected a programme with demonstrated efficacy in at least some communities, used careful strategies to adapt it to a new culture, conducted initial pilot work (a smaller controlled trial), and then extended their evaluation to a larger RCT.

Alternative evaluation designs

As noted above, there are some interventions that do not lend themselves well to evaluation by conventional methods. The Center for Evaluation Innovation discusses, for example, how complex, system-wide innovations requiring continuous adaptations may not be easily evaluated with experimental outcome evaluation frameworks (Connell, Kubisch, Schorr & Weiss, 1995; Preskill & Beer, 2012).

Advocacy interventions sometimes also fall into this category. While the need to define programme theory and outcomes still applies, it can be difficult to evaluate advocacy efforts with quasi-experimental designs and RCTs (Reisman, Gienapp & Stachowiak, 2007). Part of the problem lies in the fact that although an advocacy effort may ultimately aim to increase the protection of children from violence at some point in the future, its first set of goals may involve strengthening protective government legislation and enhancing the functioning of protective agencies. In this particular example, children and families may not see any improvements in their circumstances within the time frame of a few years. In addition, it might be difficult to collect relevant outcome data for controls (or under comparison conditions); random assignment might also be impossible.

One possible option in these circumstances is to focus initially on shorter-term community-level or system-level outcomes, with the proviso that child and family-level outcomes will be assessed at a later stage. Another option is to collect outcome data relating to children and families as a part of your evaluation with the understanding that, although changes might not be detectable in the short term, the data will serve as a baseline for future follow-up studies.

However, even if at first it seems that your intervention does not easily lend itself to the outcome evaluation methodologies described above, it is worth taking the time to fully consider this with your research partners. Many researchers have found creative ways to use rigorous study designs to evaluate the effectiveness of system- or community-level policy or advocacy interventions. Lagarde, Haines & Palmer (2007), for example, successfully conducted a series of randomized controlled evaluations of conditional cash transfers in multiple low- and middle-income countries and valuable information about the effectiveness of the large-scale public intervention was obtained.

Regardless of whether the programme or intervention that you are developing is concrete and targeted or complex and multi-layered, if your intention is to replicate it in other communities or to scale it up, then a rigorous, traditional outcome evaluation is likely your best route.

Applying this handbook to your programmes

The purpose of this handbook is to increase awareness among agencies engaged in developing and implementing interventions to reduce children's exposure to violence of the need to incorporate research and evaluation into their programming. In the current culture of evidence, funders are increasingly looking for this level of commitment so that their limited funds are spent on programmes that achieve results, and so that the field can collectively build on information about what interventions best protect children from harm.

There is a huge diversity in the size and focus of agencies wanting to develop and implement measures to improve the safety and well-being of children subject to violence. We do not assume that all agencies are equally equipped to implement a rigorous evaluation of such efforts. In-house agency evaluation and research expertise varies widely. Furthermore, the type of interventions that agencies want to put in place, and the degree to which they have been tried and tested in other areas, also differs greatly. So too does the context and setting for these interventions: across low- and middle-income countries especially, there is a wide disparity in the level of infrastructure, political stability and availability of resources and also in the nature and level of violence that affects children.

All of these factors will play a role in determining how easily your organization will be able to meet the suggestions put forward in this handbook. You will have to assess your situation and draw from the handbook accordingly. However, do not assume that a focus on difficult problems in low-resourced communities means that rigorous evaluation is not possible. As we note above, there are many examples of successful applications of very high-quality evaluation in such locations.

In presenting our recommendations, we do not prescribe a particular linear pathway to be followed as you build your programme theory, define outcomes and plan for evaluation. These efforts often occur simultaneously and iteratively as agencies seek to improve their work. Even programmes and agencies that have long histories of rigorous evaluation follow different paths to development, evaluation, improvement and expansion.

Your goal should be to advance as much as possible your own participation in the drive to create and implement evidence-based programmes to protect children from violence. Use the handbook to develop or improve your programme theory, to define measurable outcomes, to build research partnerships, to identify strengths and weaknesses in your own use of research and evaluation, and to make plans to increase your capacity in these areas. If all agencies, programme developers, researchers and funders work together, successful efforts will be adopted more widely and more quickly to the ultimate benefit of children everywhere.

References

- Aniekwe CC, Hayman R, Mdee A (2012). Academic-NGO collaboration in international development research: A reflection on the issues. Oxford: International NGO Training and Research Centre (http://www.intrac.org/data/files/resources/750/Academic-NGO-Collaboration-in-International-Development_September-2012.pdf, accessed 17 January 2014).
- Baker-Henningham H (2011). Transporting evidence-based interventions across cultures: using focus groups with teachers and parents of pre-school children to inform the implementation of the Incredible Years Teacher Training Programme in Jamaica. *Child Care Health Dev.* 37(5):649–661.
- Baker-Henningham H, Scott S, Jones K, Walker S (2012). Reducing child conduct problems and promoting social skills in a middle-income country: cluster randomised controlled trial. *Br J Psychiatry.* 201(2):101–108.
- Baker-Henningham H, Walker S, Powell C, Gardner JM (2009). A pilot study of the Incredible Years Teacher Training programme and a curriculum unit on social and emotional skills in community pre-schools in Jamaica. *Child Care Health Dev.* 35(5):624–631.
- Barr RG, Barr M, Fujiwara T, Conway J, Catherine N, Brant R (2009). Do educational materials change knowledge and behaviour about crying and shaken baby syndrome? A randomized controlled trial. *CMAJ.* 180(7):727–733.
- Barr RG, Rivara FP, Barr M, Cummings P, Taylor J, Lengua LJ et al. (2009). Effectiveness of educational materials designed to change knowledge and behaviors regarding crying and shaken-baby syndrome in mothers of newborns: a randomized, controlled trial. *Pediatrics.* 123(3):972–980. doi: 10.1542/peds.2008-0908.
- Barrera M, Castro FG (2006). A heuristic framework for the cultural adaptation of interventions. *Clinical Psychology: Science and Practice.* 13(4):311–316. doi: 10.1111/j.1468-2850.2006.00043.x
- Betancourt TS, Meyers-Ohki SE, Stevenson A, Ingabire C, Kanyanganzi F, Munyana M et al. (2011). Using mixed-methods research to adapt and evaluate a family strengthening intervention in Rwanda. *Afr J Trauma Stress.* 2(1):32–45.
- Castro FG, Barrera Jr M, Holleran Steiker LK (2010). Issues and challenges in the design of culturally adapted evidence-based interventions. *Annu Rev Clin Psychol.* 6:213–239. doi: 10.1146/annurev-clinpsy-033109-132032.
- Chaffin M, Friedrich B (2004). Evidence-based treatments in child abuse and neglect. *Child Youth Serv Rev.* 26(11):1097–1113.
- Chaffin M, Silovsky JF, Funderburk B, Valle LA, Brestan EV, Balachova T et al. (2004). Parent-child interaction therapy with physically abusive parents: Efficacy for reducing future abuse reports. *J Consult Clin Psychol.* 72(3):500–510.

- Chen H-T (2005). *Practical program evaluation*. Thousand Oaks, CA: Sage Publications Inc.
- Child Protection Monitoring and Evaluation Reference Group (2012). *Ethical principles, dilemmas and risks in collecting data on violence against children: A review of the available literature*. New York, NY: Statistics and Monitoring Section, Division of Policy and Strategy, United Nations Fund for Children.
- Cohen JA, Deblinger E, Mannarino AP, Steer R (2004). A multisite, randomized controlled trial for children with abuse-related PTSD symptoms. *J Am Acad Child Adolesc Psychiatry*. 43(4):393–402.
- Connell JP, Kubisch AC, Schorr LB, Weiss CH (eds) (1995). *New approaches to evaluating community initiatives*. Washington DC: Aspen.
- Cross TP, Jones L, Walsh W, Simone M, Kolko D, Szczepanski J et al. (2008). Evaluating Children's Advocacy Centers' response to child sexual abuse. In: *Juvenile Justice Bulletin*. Washington, DC: US.Department of Justice:1–12.
- Cuijpers P, De Graaf I, Bohlmeijer E (2005). Adapting and disseminating effective public health interventions in another country: towards a systematic approach. *The European Journal of Public Health*. 15(2):166–169.
- De Koker P, Mathews C, Zuch M, Bastien S, Mason-Jones AJ (2014). A systematic review of interventions for preventing adolescent intimate partner violence. *J Adolesc Health*. 54(1):3–13.
- Durlak JA, Weissberg RP, Dymnicki AB, Taylor RD, Schellinger KB (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Dev*. 82(1):405–432. doi: 10.1111/j.1467-8624.2010.01564.x.
- Durlak JA, Weissberg RP, Pachan M (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *Am J Community Psychol*. 45(3–4):294–309. doi: 10.1007/s10464-010-9300-6.
- Eisner M (2009). No effects in independent prevention trials: can we reject the cynical view? *J Exp Criminol*. 5(2):163–183.
- Farrelly MC, Niederdeppe J, Yarsevich J (2003). Youth tobacco prevention mass media campaigns: past, present, and future directions. *Tob Control*. 12(Suppl. 1):i35–i47.
- Filene JH, Kaminski JW, Valle LA, Cachat P (2013). Components associated with home visiting program outcomes: A meta-analysis. *Pediatrics*. 132(Suppl. 2):S100–S109. doi: 10.1542/peds.2013-1021H.
- Finkelhor D (1979). *Sexually victimized children*. New York, NY: Free Press.
- Finkelhor D, Hamby S, Turner H, Walsh AW (in press). Ethical issues in surveys about children's exposure to violence and sexual abuse. In: Cuevas CA, Rennison CM (eds). *The Wiley-Blackwell Handbook on the psychology of violence*. Oxford, UK: Wiley-Blackwell.
- Foshee VA, Bauman KE, Arriaga XB, Helms RW, Koch GG, Linder GF (1998). An evaluation of Safe Dates, an adolescent dating violence prevention program. *Am J Public Health*. 88(1):45–50.
- Funnell SC, Rogers PJ (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco, CA: Jossey-Bass.
- Hutchings J, Bywater T, Eames C, Martin P (2008). Implementing child mental health interventions in service settings: Lessons from three pragmatic randomised controlled trials in Wales. *J Child Serv*. 3(2):17–27.

- Justice Research and Statistics Association (2003). *Evaluability assessment: Examining the readiness of a program for evaluation*. Washington, DC: Office of Justice Programs, US Department of Justice (Program Evaluation Briefing Series 6; <http://www.jrsa.org/pubs/just-justice/evaluability-assessment.pdf>, accessed 7 July 2014).
- Kanof ME (2003). *Youth illicit drug use prevention: DARE long-term evaluations and federal efforts to identify effective programs*. Washington DC: General Accounting Office.
- Knerr W, Gardner F, Cluver L (2013). Improving positive parenting skills and reducing harsh and abusive parenting in low-and middle-income countries: A systematic review. *Prev Sci*. 14(4):352–363. doi: 10.1007/s11121-012-0314-1.
- Lagarde M, Haines A, Palmer N (2007). Conditional cash transfers for improving uptake of health interventions in low-and middle-income countries. *JAMA*. 298(16):1900–1910.
- Lantz PM, Jacobson PD, Warner KE, Wasserman J, Pollack HA, Berson J et al. (2000). Investing in youth tobacco control: A review of smoking prevention and control strategies. *Tob Control*. 9(1):47–63.
- MacMillan HL, Wathen CN, Barlow J, Fergusson DM, Leventhal JM, Taussig HN (2009). Interventions to prevent child maltreatment and associated impairment. *Lancet*. 373(9659):250–266. doi: 10.1016/S0140-6736(08)61708-0. Epub 2008 Dec 4.
- Nowak C, Heinrichs N (2008). A comprehensive meta-analysis of Triple P – Positive Parenting Program using hierarchical linear modeling: Effectiveness and moderating variables. *Clin Child Fam Psychol Review*. 11(3):114–144. doi: 10.1007/s10567-008-0033-0.
- Patton MQ (1997). *Utilization-focused evaluation*, 3rd ed. Thousand Oaks, CA: Sage Publications Inc.
- Petrosino A, Turpin-Petrosino C, Buehler J (2003). Scared Straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review of the randomized experimental evidence. *Ann Am Acad Pol Soc Sci*. 589(1):41–62.
- PREPARE: Promoting sexual and reproductive health among adolescents in southern and eastern Africa [website]. Bergen, University of Bergen (<http://prepare.b.uib.no/>, accessed 10 July 2014).
- Preskill H, Beer T (2012). *Evaluating social innovation*. Washington, DC: Center for Evaluation Innovation (<http://www.evaluationinnovation.org/sites/default/files/EvaluatingSocialInnovation.pdf>, accessed 8 July 2014).
- Rahman A, Iqbal Z, Roberts C, Husain N (2009). Cluster randomized trial of a parent-based intervention to support early development of children in a low-income country. *Child Care Health Dev*. 35(1):56–62. doi: 10.1111/j.1365-2214.2008.00897.x.
- Reisman J, Gienapp A, Stachowiak S (2007). *A guide to measuring advocacy and policy*. Baltimore, MD: Annie E. Casey Foundation (<http://www.aecf.org/resources/a-guide-to-measuring-advocacy-and-policy/>, accessed 8 July 2014.)
- Ringold DJ (2002). Boomerang effects in response to public health interventions: Some unintended consequences in the alcoholic beverage market. *Journal of Consumer Policy*. 25(1), 27–63.
- Rossi PH, Lipsey MW, Freeman HE (2004). *Evaluation: A systematic approach*, 7th ed. Thousand Oaks, CA: Sage Publications Inc.

- Sanders MR, Turner KMT, Markie-Dadds C (2002). The development and dissemination of the Triple P – Positive Parenting Program: A multilevel, evidence-based system of parenting and family support. *Prev Sci.* 3(3):173–189.
- Stallard P (1996). The role and use of consumer satisfaction surveys in mental health services. *J Ment Health.* 5(4):333–348.
- Swearer SM, Espelage DL, Vaillancourt T, Hymel S (2010). What can be done about school bullying? Linking research to educational practice. *Educational Researcher*, 39(1):38–47.
- Tharp AT (2012). Dating Matters™: The next generation of teen dating violence prevention. *Prev Sci.* 13(4):398–401. doi: 10.1007/s11121-012-0307-0.
- W.K. Kellogg Foundation (1998). W.K. Kellogg Foundation evaluation handbook. Battle Creek, MI: W.K. Kellogg Foundation (<http://www.wkkf.org/resource-directory/resource/2010/w-k-kellogg-foundation-evaluation-handbook>, accessed 10 July 2014).
- Werch CE, Owen DM (2002). Iatrogenic effects of alcohol and drug prevention programs. *J Stud Alcohol.* 63(5):581–590.
- Wessels I, Mikton C, Ward CL, Kilbane T, Alves R, Campello G et al. (2013). Preventing violence: Evaluating outcomes of parenting programmes. Geneva: World Health Organization (http://www.who.int/violence_injury_prevention/publications/violence/parenting_evaluations/en/, accessed 10 July 2014).
- Wholey JS, Hatry HP, Newcomer KE (eds) (2004). *Handbook of practical program evaluation*, 2nd ed. San Francisco, CA: Jossey-Bass.
- Wilson P, Rush R, Hussey S, Puckering C, Sim F, Allely CS et al. C (2012). How evidence-based is an 'evidence-based parenting program'? A PRISMA systematic review and meta-analysis of Triple P. *BMC Med.* 10(1):130. doi: 10.1186/1741-7015-10-130.
- Wolfe DA, Crooks C, Jaffe P, Chiodo D, Hughes R, Ellis W et al. (2009). A school-based program to prevent adolescent dating violence: A cluster randomized trial. *Arch Pediatr Adolesc Med.* 163(8):692–699. doi: 10.1001/archpediatrics.2009.69.

APPENDIX

Evaluation resources

This Appendix provides an array of resources to applicants including helpful websites and publications on evaluation, available libraries and ways to access scholarly journals and search engines, and compendiums of validated outcome measures and scales relevant to this field.

Text books on evaluation

Chen H-T (2005). Practical program evaluation. Thousand Oaks, CA: Sage Publications, Inc.

Funnell SC, Rogers PJ (2011). Purposeful program theory: Effective use of theories of change and logic models. San Francisco, CA: Jossey-Bass.

Patton MQ (1997). Utilization-focused evaluation, 3rd ed. Thousand Oaks, CA: Sage Publications, Inc.

Rossi PH, Lipsey MW, Freeman HE (2004). Evaluation: A systematic approach, 7th ed. Thousand Oaks, CA: Sage Publications, Inc.

Wholey JS, Hatry HP, Newcomer KE, editors (2004). Handbook of practical program evaluation, 2nd ed. San Francisco, CA: Jossey-Bass.

Organizations with online information on programme theory, outcome measurement and evaluation

The American Evaluation Association (AEA) (www.eval.org).

The US Centers for Disease Control (CDC) (www.cdc.gov). (See in particular the following evaluation resources: <http://www.cdc.gov/eval/resources/index.htm> and <http://www.cdc.gov/eval/framework/index.htm>).

The National Collaborating Center for Methods and Tools (<http://www.nccmt.ca/eiph/define-eng.html>).

The National Resource Center for Community-Based Child Abuse Prevention (CBCAP) (<http://friendsnrc.org/evaluation-toolkit>).

Perform Well (<http://www.performwell.org/>).

The Center for Evaluation Innovation (www.evaluationinnovation.org).

The University of Wisconsin Cooperative Extension offers a free online course on creating logic models (available at: <http://www.uwex.edu/ces/lmcourse/>).

UNICEF (Evaluation and Good Practices) (<http://www.unicef.org/evaluation/index.php>).

The W.K. Kellogg Foundation (www.wkkf.org). (See for example their guide on logic model development available at: <http://www.wkkf.org/knowledge-center/resources/2006/02/WK-Kellogg-Foundation-Logic-Model-Development-Guide.aspx> and their guide on evaluations available at: <http://www.wkkf.org/knowledge-center/resources/2010/W-K-Kellogg-Foundation-Evaluation-Handbook.aspx>).

The World Health Organization (WHO). (See for example their parenting outcome evaluation toolkit available at: http://www.who.int/iris/bitstream/10665/85994/1/9789241505956_eng.pdf and a web-appendix with numerous additional resources available at: http://www.who.int/violence_injury_prevention/publications/violence/parenting_programmes_webappendix.pdf).

Online resources on budgeting for evaluation

W.K. Kellogg Foundation (<http://www.wkkf.org/resource-directory/resource/2010/w-k-kellogg-foundation-evaluation-handbook>).

Western Michigan University Evaluation Center (http://www.wmich.edu/evalctr/archive_checklists/evaluationbudgets.pdf).

Pell Institute Evaluation Toolkit (<http://toolkit.pellinstitute.org/evaluation-guide/plan-budget/develop-a-budget/>).

BetterEvaluation (http://betterevaluation.org/evaluation-options/develop_eval_budget_matrix).

Websites and reports providing information on evidence-supported programmes and practices

Centers for Disease Control (<http://vetoviolence.cdc.gov/>).

National Registry of Evidence-Based Programs and Practices (<http://www.nrepp.samhsa.gov/>).

Blueprints for Violence Prevention (<http://www.colorado.edu/cspv/blueprints/>).

Kauffman Best Practices Project (<http://www.chadwickcenter.org/Documents/Kaufman%20Report/ChildHosp-NCTA brochure.pdf>).

Handbook of Injury and Violence Prevention (pdfs of book sections available at: <http://link.springer.com/book/10.1007/b136518/page/1>).

Guide to Community Preventive Services (<http://www.thecommunityguide.org/index.html>).

California Evidence-Based Clearinghouse (<http://www.cebc4cw.org/>).

What Works Clearinghouse (<http://ies.ed.gov/ncee/wwc/>).

Find Youth Info (<http://www.findyouthinfo.gov/>).

Promising Practices Network for Children, Families, and Communities (<http://www.promisingpractices.net/>).

World Health Organization's Violence Prevention: The Evidence http://www.who.int/violence_injury_prevention/violence/4th_milestones_meeting/publications/en/.

Safe Start Center (<http://www.safestartcenter.org/research/>).

The Campbell Collaboration (<http://www.campbellcollaboration.org/index.html>).

Saunders, B.E., Berliner, L., & Hanson, R.F. (Eds.). (2004). Child Physical and Sexual Abuse: Guidelines for Treatment (Revised Report: April 26, 2004). Charleston, SC: National Crime

Victims Research and Treatment Center. (<http://tfcbt.musc.edu/resources/pdfs/Guidelines%20Final.pdf>).

Child victimization research centres and organizations

Centers for Disease Control and Prevention (CDC), GA, USA (<http://www.cdc.gov/violenceprevention/childmaltreatment/>).

Crimes Against Children Research Center (CCRC), NH, USA (www.unh.edu/ccrc)

International Society for the Prevention of Child Abuse and Neglect (ISPCAN), CO, USA (www.ispcan.org).

National Crime Victims Center, Medical University of South Carolina (MUSC), SC, USA (<http://colleges.musc.edu/ncvc>).

National Society for the Prevention of Cruelty to Children, United Kingdom, (<http://www.nspcc.org.uk/>).

The Child Protection Research Centre, a collaboration between the University of Edinburgh and the National Society for the Prevention of Cruelty to Children (NSPCC), United Kingdom (<http://www.childprotection.ed.ac.uk/>).

The Kempe Center for the Prevention and Treatment of Child Abuse and Neglect, CO, USA (<http://www.ucdenver.edu/academics/colleges/medicalschoo/departments/pediatrics/subs/can/Pages/ChildAbuseNeglect.aspx>).

World Health Organization (Dept. of Violence Prevention), Geneva, Switzerland (http://www.who.int/violence_injury_prevention/violence/child/en/index.html).

Trial registration information

There are increasing calls for all controlled trials to be registered in clinical trial databases, so that the results of all studies (even unpublished ones) can be reviewed and used by researchers. Existing databases for clinical trials relating to violence prevention include:

Pan-African Clinical Trials Registry (www.pactr.org).

Violence Prevention Evidence Base (<http://www.preventviolence.info/>).

Violence Prevention Trials Registry (http://www.preventviolence.info/trials_search.aspx).

WHO International Clinical Trials Registry Platform Central Database (www.who.int/trialsearch/).

Clinical Trials.gov (www.clinicaltrials.gov).



**World Health
Organization**

Department for Management of Noncommunicable Diseases, Disability,
Violence and Injury Prevention (NVI)

World Health Organization

20 Avenue Appia

CH-1211 Geneva 27

Switzerland

Tel +41-22-791-2064

Fax +41-22-791-4489

www.who.int/violence_injury-prevention

violenceprevention@who.int



Crimes Against Children Research Center

University of New Hampshire

Lisa M. Jones, Ph.D. (lisa.jones@unh.edu)

United States

Tel +1-603-862-2515

www.unh.edu/ccrc

ISBN 978 92 4 150788 2



9 789241 507882