Selecting outcomes for intimate partner violence intervention trials: overview and recommendations

Dr Lorna Jane O'Doherty[a,f] (corresponding author), Professor Harriet MacMillan[b], Professor Gene Feder[c], Professor Angela Taft[d], Professor Ann Taket[e], Professor Kelsey Hegarty[f]

[a]Centre for Research in Psychology, Behaviour and Achievement, Coventry University, Priory Street, Coventry, UK Tel: +44(0)2477659521 lornajaneodoherty@hotmail.com

[b]Offord Centre for Child Studies, Department of Psychiatry and Behavioural Neurosciences and of Pediatrics, McMaster University, Wilcox Building, 1280 Main St, Hamilton, ON L8S 4K1, Canada
Tel: 001(0)9055212100 macmilnh@mcmaster.ca

[c]Academic Unit of Primary Health Care, University of Bristol, 25 Belgrave Road, Clifton, Bristol BS8 2AA UK Tel: +44(0)1173313865 gene.feder@bristol.ac.uk

[d]Judith Lumley Centre, La Trobe University, 215 Franklin St, Vic 3000, Australia, Tel +61 (0)394798809 a.taft@latrobe.edu.au

[e]School of Health & Social Development, Faculty of Health, Deakin University, Melbourne Burwood Campus, 221 Burwood Highway, Burwood, VIC 3125, Australia
ann.taket@deakin.edu.au

[f]Department of General Practice, The University of Melbourne, 200 Berkeley Street, Carlton, Vic 3031, Australia, Tel: +61 (0)383444992 k.hegarty@unimelb.edu.au

ABSTRACT

Intimate partner violence (IPV) is endemic in societies around the world and detrimental to women's wellbeing. Abused women are frequent users of health services. Despite the recent World Health Organization guidelines on IPV and sexual violence, we need more evidence on effective responses to women in health care settings. Developing robust evidence with potential to inform policy and clinical practice requires greater clarity and consistency across studies in the selection and use of outcomes to evaluate interventions. Drawing on systematic reviews and individual trials aimed at reducing abuse and improving women's health, we discuss critical issues in respect of outcomes. We discuss primary, secondary, intermediate and proxy outcomes and measures used to evaluate interventions for women who experience IPV. We offer recommendations about which outcomes to assess and approaches to doing so within the context of trials in health care settings.

Keywords: Intimate partner violence; Trials; Outcomes; Evaluation; Measurement

Introduction

Intimate partner violence (IPV) is defined as any behavior within an intimate relationship that causes physical, psychological or sexual harm to those in the relationship (Krug, Mercy, Dahlberg, & Zwi, 2002). The latest data from the World Health Organization (WHO) suggest that one in three women experience physical or sexual violence in their lifetime (World Health Organization, 2013a). Violence against women by male partners is a significant contributor to illness, disability and death of women around the world (Black, 2011; Heise, Ellsberg, & Gottmoeller, 2002).

Exposure to violence by partners leads to frequent use of health services suggesting that health care settings offer opportunities to support women and intervene in the cycle of violence (Bonomi, Anderson, Rivara, & Thompson, 2009). Although there is some evidence for responding to IPV within health care settings as identified in the recent WHO Guidelines (World Health Organization, 2013b), there is still an urgent need for rigorous trials to determine the effectiveness of interventions in low-, middle- and high-income countries. Interventions for IPV in health care settings mainly comprise those that screen women, offer support/advocacy or involve a therapeutic intervention (MacMillan et al., 2001). Generally, these programs aim to safeguard women from harm, manage symptoms, improve health outcomes, encourage safe communication with others when in an abusive relationship, increase women's confidence to seek help, and improve social networks (Ford-Gilboe, Merritt-Gray, Varcoe, & Wuest, 2011).

Synthesis of different studies has been conducted in respect of screening-only interventions (Feder et al., 2009; Nelson, Bougatsos, & Blazina, 2012; Taft et al., 2013), advocacy (Ramsay et al., 2009) and interventions (with or without screening) to reduce IPV in primary care (Bair-Merritt et al., 2014) and against pregnant women (Jahanfar, Janssen, Howard, & Dowswell, 2013). The limited number

of trials overall, heterogeneity of interventions (e.g. treatment intensity) (Ramsay et al., 2009), and barriers to achieving high-quality evidence (e.g. difficulties in minimizing the risk of bias) impede the emergence of better quality evidence of effective IPV interventions (O'Doherty et al., 2014). An additional barrier concerns the inconsistent selection of effect measures that might allow syntheses of evidence if there was some consistency across trials.

The benefits of meta-analysis for improving evidence include increasing power and precision, answering questions not posed by individual studies, and opportunities to settle controversies arising from conflicting findings (Higgins & Green, 2011). However evidence from meta-analyses is particularly lacking in this field, hampering the process of translating research evidence into what can be safely and effectively delivered in clinical practice. Meta-analyses have been particularly challenged by the lack of clarity and consistency in effect measures (O'Doherty et al., 2014; Ramsay et al., 2009). Robustness of evidence can be enhanced where evaluations are based on better informed decisions around selection of outcomes, approaches to measurement and timeframes. The complexity of IPV trials means that evaluation can involve multiple dimensions, most typical being abuse, social, clinical, wellbeing and service use outcomes. The multiplicity of areas of perceived importance leads to significant heterogeneity in outcomes across trials (O'Doherty et al., 2014; Ramsay et al., 2009). Within certain domains, a large number of tools is available which increases difficulties with combining results. Across the ten trials included in a review of advocacy interventions (Ramsay et al., 2009), eight measured abuse using six different scales (or subscales) and a single item question. Another issue is the lack of measurement in certain areas e.g. those associated with children and parenting and harm. In a review of IPV interventions for pregnant women, Jahanfar and colleagues (p.8) indicated that “a serious problem in this review was the lack of consistency in, and the limited range of outcomes reported, and the varied way that outcomes such as depression or experience of violence were measured.” Ramsay and colleagues (2009) called for

debate over which outcomes should be measured, which outcomes could even be expected to change in response to an intervention and a more standardized approach to measurement.

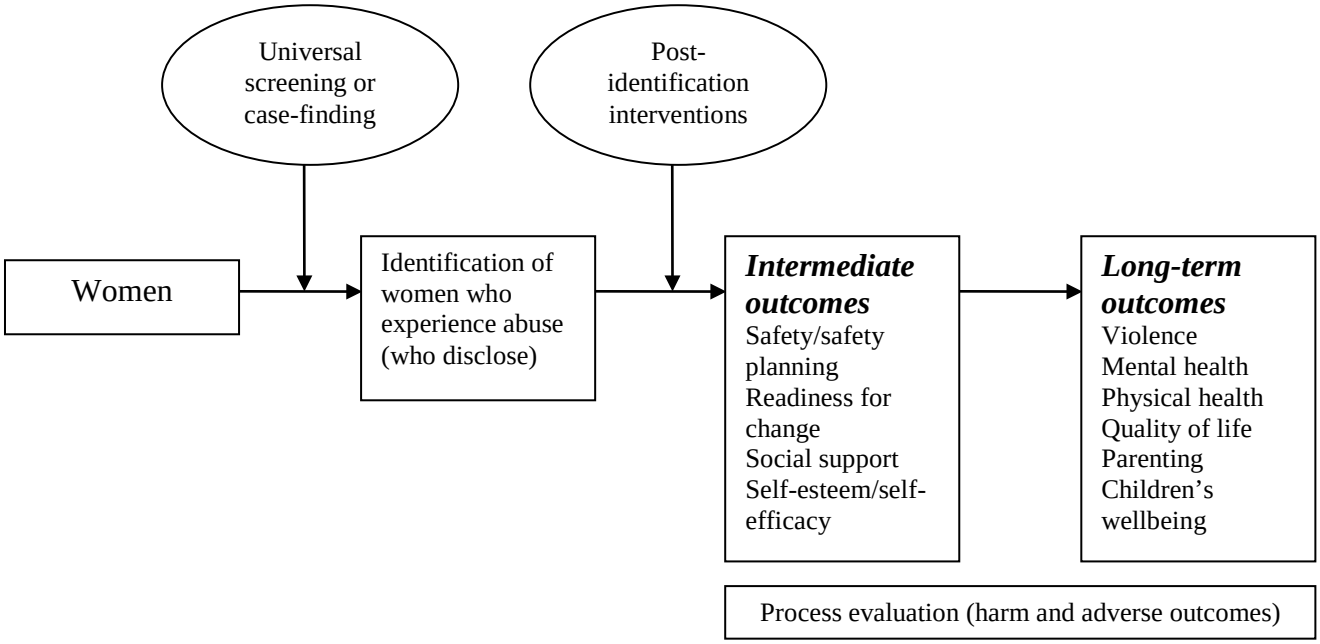
Thus, there are a number of critical questions about outcome selection. What outcomes are meaningful for survivors of IPV in health care contexts? Which outcomes should be primary outcomes? Which can be nominated as secondary outcomes? What intermediate and proxy outcomes are relevant? We will discuss critical issues in outcome selection, and offer recommendations for future trials in health care settings.

Approach

We will draw on an analytic model around the delivery of IPV interventions for women in health care settings (MacMillan et al., 2001) as a framework to discuss the different outcomes in IPV trials (Figure 1). The objective is not to focus on results *per se*, but rather to examine how effectiveness is being assessed. In discussing outcome selection, we draw on systematic reviews where they are available (Bair-Merritt et al., 2014; Feder et al., 2009; Jahanfar et al., 2013; Ramsay et al., 2009; Taft et al., 2013). Otherwise, we refer to individual trials mainly in health care settings. We will focus on outcomes associated with interventions that involve direct interactions with women experiencing IPV. These include screening interventions and those offered post-identification, such as advocacy, support and psychological interventions. Figure 1 captures hypothesized pathways to improved outcomes for women experiencing IPV. It highlights identification as the main outcome of interest in screening trials and a range of intermediate and long-term outcomes of relevance all types of IPV interventions.

Figure 1 here

Figure 1 Interventions for IPV in health care setting and outcomes. Adapted from MacMillan et al. (2001)



Critical issues in outcome selection and measurement

1. Primary and secondary outcomes

Conclusions about the effects of an intervention will be based largely on the primary outcomes, as trials should be powered on those outcomes, although low power can be mitigated by pooling in meta-analyses. Primary outcomes are the outcomes that are essential for decision-making. It is recommended that there are no more than three, that surrogate or interim outcomes are not included and there is at least one desirable and one undesirable outcome (to assess adverse as well as beneficial effects) (Higgins & Green, 2011). Main outcomes that were not selected as primary outcomes can be listed as secondary outcomes. Secondary outcomes may also include measures that are less important than clinical endpoints in informing decisions, but which may be helpful in explaining the effect or determining intervention integrity (Higgins & Green, 2011). Trials are not necessarily powered to detect important differences in secondary outcome measures although, like primary outcome measures, they can be pooled across trials in meta-analyses.

2. Issues specific to screening trials

Screening interventions aim to identify abused women in order to offer interventions leading to beneficial outcomes. Although there have been calls for a shift in focus from screening to interventions that can be offered to women once identified (Wathen & MacMillan, 2012), the need to evaluate screening against active case-finding remains (O'Doherty et al., 2014). Furthermore, as screening plus referral is the most commonly implemented response to abuse, therapeutic interventions need to be separately evaluated – versus usual care, case-finding and screening. Variables such as identification rates and referral are common to screening trials and integral to the short-term evaluation of screening interventions (Taft et al., 2013). Screening reviews have demonstrated the wide range of screening tools in use (Feder et al., 2009; Nelson et al., 2012). One problem is rates of identification tending to be lower when undertaken directly by clinicians compared to anonymous disclosure (e.g. self-completion methods such as written or computerized surveys) regardless of the tools used (O'Doherty et al., 2014). This leads to difficulties in statistically combining results derived using varied approaches (MacMillan et al., 2006). Some studies use disclosure or discussion about safety as a proxy for identification (Taft et al., 2012), rather than documentation in records as the latter varies across clinicians and is likely to be an underestimate of identification (O'Doherty et al., 2014).

Trials often aim to measure referrals. However, screening studies need to differentiate the provision of lists of resources/services from referral, which involves a more formal process of linking women with other clinical and specialist support services. In the IRIS trial, testing a domestic violence (DV) training and support intervention to general practices, the primary outcome was a general practitioner referral (of a patient disclosing abuse) to specialist DV agencies (Feder et al., 2011). The subsequent uptake of services may be measured through using a set of survey questions or contacting the relevant services to determine uptake rates but rates of missing data tend to be high (O'Doherty et al.,

2014). Moreover, there is the potential for a mismatch between referrals recorded in the medical record and referrals recorded as received in the target service (Feder et al., 2011). In a review of screening trials, Taft and colleagues hypothesized that screening would firstly improve identification rates, the provision of information, and referral to support services (Taft et al., 2013) and that these would (in the longer-term) lead to reductions in abuse and improvement in health and wellbeing. However, of the 11 trials, only two measured abuse (Koziol-McLain et al., 2010; MacMillan et al., 2009) and one measured health (MacMillan et al., 2009). This indicates a need to evaluate screening interventions in respect of the long-term impacts on women's health and wellbeing (O'Doherty et al., 2014), a topic addressed in Section 4 in respect of all types of IPV intervention.

3. Intermediate outcomes

The Medical Research Council emphasized the importance of modelling process and outcome in a trial's development phase (Craig et al., 2008), encouraging researchers to hypothesize about how any impact of exposure on outcome might be achieved. An intermediate outcome (or mediating variable) refers to any factor that represents a step in the causal pathway between the exposure and outcome (Gunasekara, Carter, & Blakely, 2008). Intermediate outcomes should be included when evaluating IPV interventions given the complexity generated by their multifaceted nature and tendency to involve a multi-step pathway to change.

Measuring intermediate outcomes implies collecting data on an occasion prior to the assessment of final trial outcomes in order to capture the 'middle ground' and temporal aspects of the change pathway. This may be illustrated in a trial that evaluated brief counselling by primary care doctors; women's readiness for change and reports of doctor enquiry about safety were measured at 6 months ahead of the final outcomes, quality of life and mental health, at one year (Hegarty et al., 2013). Other studies have gathered data relevant to understanding causal pathways retrospectively. Klevens and colleagues, in their trial of computerized enquiry plus provision of resources (Klevens et al.,

2012), asked women at one year follow-up if they had contacted any of the listed services or an IPV agency alongside assessing the primary outcome (physical and mental health on SF-12). The costs associated with recall bias and missing data need to be weighed against costs of undertaking interim data collection. Despite studies including intermediate outcomes, generally, there are few reports of causal chains or models (Hegarty, O'Doherty, Gunn, Pierce, & Taft, 2008) This suggests that opportunities for disseminating conceptual knowledge as regards complex IPV interventions are being missed.

Intermediate outcomes should not be confused with confounding variables and as such must be treated differently in analyses (Seuc, Peregoudov, Betran, & Gulmezoglu, 2013). Confounders are extraneous variables that are expected to correlate with the exposure and outcome. IPV interventions are particularly susceptible to contextual factors given the sensitivity of the issue. Effective randomization should deal with confounders, and minimize the need to include them. Distinguishing confounders from mediators of the intervention's effect requires articulation of an explicit causal pathway in the design of the trial.

Below we discuss examples of intermediate outcomes relevant once women have been identified (Figure 1). Typical outcomes include safety and safety planning, readiness for change, social support, and self-esteem and self-efficacy. Intermediate outcomes of interest may be designated as a trial's secondary outcomes. Another approach might be to include primary outcome measures at intermediate stages as the secondary outcomes. This approach has been taken in the PATH trial with the primary depression and psychological distress questionnaires at 4 and 8 months, in addition to the final 12 months follow-up (Brierley et al., 2013). One of the key issues here is distinguishing intermediate and long-term outcomes and ensuring *a priori* decisions about the role of each outcome.

Also intermediate outcomes are assumed to be related to long-term outcomes, which may not always be the case.

3.1 Safety and safety planning

While women have no control over their partner's violence, it may be possible for them to behave in ways that promote their safety and wellbeing, particularly during times of crisis. Increasing women's awareness of how to stay safe has been central to a number of interventions (Hegarty et al., 2013; McFarlane, Groff, O'Brien, & Watson, 2006; McFarlane, Soeken, & Wiist, 2000; Parker, McFarlane, Soeken, Silva, & Reel, 1999; Tiwari et al., 2005). Despite the appeal of safety indicators as intermediate or proxy variables and as primary trial endpoints, few studies have included them (Ramsay et al., 2009; Taft et al., 2013). Assessing women's actual safety behaviors may represent a more valid measure of safety than simply enquiring about the existence of safety plans. The notion of a safety plan may be too abstract, particularly among women who are in the early stages of recognizing the problem. The Safety Behavior Checklist (McFarlane, Parker, Soeken, Silva, & Reel, 1998) which measures safety behaviors (e.g. hiding documents, establishing a signal) has been the tool of choice in a number of trials (Gillum, Sun, & Woods, 2009; Hegarty et al., 2013; Tiwari, Fong, et al., 2012). However, its use is challenged by difficulties in identifying a meaningful timeframe at baseline and follow-up, and the use of the 'not applicable' category. Once a woman has enacted a safety behavior, her circumstances might change and 12 months later there is no need for her to continue to have a bag packed or to have an escape plan. Further, it is not clear whether enacting safety behaviors is linked to reduction in violence or improved health and wellbeing. Visual analogue scales have also been used to assess women's level of safety at home (Hegarty et al., 2013) but have not been validated. It is strongly advised that tools for assessing safety behaviors and safety plans are piloted in advance of implementing a trial, particularly if being used in populations that are different to the original development population. It is worth highlighting here, in respect of 'proxy' variables like safety and service use, that post-intervention changes may be associated with both

'positive' and 'negative' health outcomes for abused women and require careful interpretation. For instance, increased refuge/shelter usage may reflect proactive behavior on the behalf of abused women but it may also reflect an escalation of violence that has led to the women needing to seek safety (Ramsay et al., 2009).

3.2 Readiness for change

Reviews have highlighted the importance of tailoring health care and service responses to IPV according to the individual woman (Reisenhofer & Taft, 2013; Zink, Elder, Jacobson, & Klostermann, 2004). Understanding a woman's readiness for action may also help prevent inappropriate health care responses such as encouraging a woman to 'just leave', when this may result in increased stress, subject her to retaliatory violence from the partner, or other negative outcomes (Campbell, Webster, & Koziol-McLain, 2003). In particular, interventions need to take account of the complex decision-making and processes of change for women experiencing IPV, and the 'turning points' in the trajectory (Chang et al., 2010). In the Psychosocial Readiness Model, readiness is described as a continuum with a balance of internal and external factors determining how the woman moves from maintaining the status quo through to a desire for action or change (Cluss et al., 2006). This model suggests there are three internal factors: *awareness* is the woman's recognition that what she is experiencing is abuse; *self-efficacy* is the woman's belief that she is able to achieve difficult tasks, or cope with adversity; and *perceived support* describes the woman's sense that she is supported by those in her environment. Hegarty and colleagues (2013) designed their intervention such that family doctors were trained to assess women's readiness for change in order to decide the most appropriate 'brief counselling' technique to deploy. Other interventions have focused on the use of Dutton's empowerment model (Dutton, 1992) to increase women's self-esteem and ability to enact safety behaviors (Tiwari et al., 2005). Whilst there is a relatively sound theoretical base for the concept of readiness for change and it has been clearly described by

women in qualitative studies, measuring it has proved challenging. The Domestic Violence Survivor Assessment (Dienemann, Campbell, Landenburger, & Curry, 2002) describes the cumulative process of women's decision-making across the IPV trajectory. However, generally there is a paucity of validated tools for applying in trials. The inclusion of women who have left their partners in some intervention trials complicates this outcome even further. Therefore, including readiness to change among main trial outcomes may be problematic. Baseline levels of variables like readiness for change could be used for stratification in randomisation (Ramsay et al., 2009). This is important because characteristics of survivors (e.g. those still in relationships versus those who are not) and the context (e.g. nature of IPV, cultural factors) can influence how intervention works.

3.3 Social support

Social support from both informal and formal sources improves abused women's mental health (Coker et al., 2002), willingness and ability to seek help and their subsequent capacity to remain safe (Liang, Goodman, Tummala-Narra, & Weintraub, 2005). Beeble and colleagues explored the complex role of social support, using the Adult Social Support Questionnaire (Bogat, Chin, Sabbath, & Schwartz, 1983), on women's wellbeing and found evidence of main, mediating, and moderating effects of social support (Beeble, Bybee, Sullivan, & Adams, 2009). They reported a positive association between social support and quality of life, and a negative association with depression; and social support partially explained the effect of baseline levels of and subsequent changes in physical abuse on quality of life and depression over time. The buffering effects of social support were strongest at lower levels of abuse. Thus, influencing abused women's perception (Cluss et al., 2006) and experience of support would appear to be a worthwhile intermediate outcome of a health care intervention (Hegarty et al., 2013; Taft et al., 2011; Tiwari et al., 2010; Tiwari, Yuk, et al., 2012), whilst also being a desirable long-term endpoint.

Although the measurement of social support is less problematic than safety or readiness for change, it is not usually measured in IPV trials. Being a generic variable, the choice of measurement tools far exceeds that seen for outcomes that are specific to IPV. Instruments for assessing social support have mainly focused on informal social support. Sullivan et al. (2002; 1992) assessed social support using the Adults Social Support Questionnaire. The Interpersonal Support Evaluation List (Cohen & Williamson, 1988) was used in a number of studies (Constantino, Kim, & Crane, 2005; Mancoske, Standifer, & Cauley, 1994; Suvak, Taft, Goodman, & Dutton, 2013) including as a secondary outcome in Tiwari and colleagues' (2010) trial of a 12-week advocacy intervention comprising empowerment and telephone social support. Formal support by health care professionals has been hypothesized to be instrumental in achieving wellbeing for women (Hegarty et al., 2013).The Psychosocial Readiness model posits that perceived support is one of three internal factors which can exert either a positive or a negative influence on women's readiness to change (Cluss et al., 2006). Thus support – both informal by family and friends and formal – could be seen as a key intermediate outcome in the pathway to safety and healing (Beeble et al., 2009). In regards to the potential benefits of IPV interventions for social support in the long-term, one study (from a community setting) observed the emergence of an effect after three years (Bybee & Sullivan, 2005).

3.4 Self-esteem and self-efficacy

Lowered self-esteem is one potential outcome of IPV (Kirkwood, 1993) and thus may be selected as an endpoint for change in the short- or long-term. Also, self-esteem and related constructs (e.g. self-efficacy) may serve as resilience factors and protect women with a history of IPV from negative sequelae such as post-traumatic stress disorder (PTSD) or depression (Bradley, 2005; Reisenhofer & Taft, 2013). Self-efficacy can be described as the 'personal strength' which allows women

experiencing IPV to work towards change (Chang et al., 2006). Thus, these constructs are important as potential moderators (i.e. influencing the strength/direction of relations) between exposure and outcomes. Women with a stronger sense of self-efficacy will be more likely to attempt change and persevere in change-making behaviors in the face of challenge or obstruction. The Rosenberg Self-Esteem Scale (Rosenberg, 1965) was the most frequently used instrument in the trials examined (Arinero & Crespo, 2004; Hegarty et al., 2013; Kubany, Hill, & Owens, 2003; Edward S. Kubany et al., 2004; Labrador, Fernandez-Velasco, & Rincon, 2006). The Index of Self-Esteem (Hudson, 1982) was used in a trial of two types of social work counselling services (Mancoske et al., 1994). The related concept of self-efficacy has been measured using the Generalized Self-Efficacy Scale (Schwarzer & Jerusalem, 1995), Fry's Global and Domain-Specific Efficacy Scale (Fry & Barker, 2002) and the Self-Efficacy Scale (Scherer, 1982) while Sullivan and colleagues developed their own measure (Sullivan, Tan, Basta, Rumptz, & Davidson, 1992).

4. Trial endpoints & long-term outcomes

In this section, we discuss potential primary outcomes in IPV trials. Once the primary outcomes have been selected, the timing of evaluation needs to be specified. Time is all important in regards to trial endpoints, given that some aspects of interventions are best suited to immediate evaluation, whilst other aspects will benefit from evaluation in the medium- to long-term. For example, the review by Taft and colleagues (2013) shows how screening-only trials tend to collect endpoint data on identification and referral rates immediately after an intervention, whilst data on outcomes such as recurrence of violence and women's health and quality of life have the potential to provide valuable information when collected in the medium- to long-term. Of course, what constitutes 'long-term' varies across trials, as does the frequency and number of data collection points. Ramsay and colleagues (2009) defined short-term follow-up as up to 12 months, medium-term as from12 to 24 months, and long-term follow-up as more than two years Thus, being specific about the number of months at which

final data on primary outcomes are collected, along with whether the anchor for those time points is at baseline, at the beginning of an intervention (if different) or cessation of delivery, is necessary for accurate interpretation. Gathering outcome data in the long-term is essential since some effects are likely to attenuate over time whilst others may not emerge until sometime after an intervention has ended (Ramsay et al., 2009). As another example, Hegarty and colleagues (2013) presented final outcomes at 12 months (after initiation of intervention) and 'long-term' outcomes at 24 months (as yet unpublished).

The implications of the planned approach to administration (e.g. self-completed surveys, telephone, in-person) for the instrument's validity need to be considered. Furthermore, decisions about administration need to be made in the context of the potential risks to the safety and wellbeing of participants posed by the data collection process (Valpied and colleagues, this issue). This is particularly relevant when conducting a long-term evaluation, where the benefits of minimizing attrition rates need to be weighed against potential risks of retention. Harm is further discussed in Section 5. As set out in Figure 1, the outcomes of greatest interest as trial endpoints and in the long-term for decision-making about interventions in this field include violence, mental and physical health, quality of life and outcomes related to parenting and children.

4.1 Intimate partner violence

Many interventions with survivors of IPV aim to prevent further exposure to violence in the long term, as well as to improve health and quality of life outcomes as we discuss below. With the exception of screening trials (O'Doherty et al., 2014), the majority of intervention trials measure violence, increasingly treating it as a secondary outcome, also discussed below. The Conflict Tactics Scales (CTS and CTS-R) have been widely used to measure IPV in trials (Bair-Merritt et al., 2010; Kiely, El-Mohandes, El-Khorazaty, & Gantz, 2010; Tiwari et al., 2010; Tiwari et al., 2005; Zlotnick,

Capezza, & Parker, 2011) particularly in the USA, where the scales were developed. In other parts of the world, the Composite Abuse Scale (CAS) (Hegarty, Bush, & Sheehan, 2005) is increasingly used to measure violence. The CAS was selected as the criterion standard in an investigation of IPV prevalence and screening performance (Wathen, Jamieson, & MacMillan, 2008) and was the violence outcome measure in the two trials that evaluated screening interventions (Koziol-McLain et al., 2010; MacMillan et al., 2009). The CAS was one of two primary outcomes in Taft et al.'s (2009) Australian advocacy trial of 'mentor mothers,' and was a secondary outcome measure for the WEAVE (Hegarty et al., 2013) and PATH (Brierley et al., 2013) trials. The Partner Abuse Scale, an adaptation of the Index of Spouse Abuse, was selected as the measure in a trial of primary care-based counselling (Gillum et al., 2009). A related measure is the Danger Assessment Scale (Campbell, 1986; Campbell, 1995), which assesses severity and frequency of IPV by presenting the woman with a calendar of the past year and uses a weighted scoring system to count yes/no responses of risk factors associated with intimate partner homicide. In combination with the Severity of Violence Against Women Scale, it was used by McFarlane and colleagues in 2006 in their comparison of advocacy-based safety-promoting nurse case management with a referral card that listed a safety plan and IPV services for abused women. For a review of tools used to measure IPV, see Feder and colleagues (2009) and the CDC Compendium (Thompson, Basile, Hertz, & Sitterle, 2006).

It is not so much the range of possible tools in use as the varied follow-up periods and inconsistent reporting of data that presents the greatest difficulties in pooling violence outcomes s across trials. In Jahanfar and colleagues' (2013) review, prevention of violence during and up to one year after pregnancy and a reduction of episodes of violence were selected as primary outcomes. Their analysis was limited by there being only one trial (Kiely et al., 2010) that reported episodes of IPV during pregnancy. Further, while four trials (Curry, Durham, Bullock, Bloom, & Davis, 2006; Kiely et al., 2010; Tiwari et al., 2005; Zlotnick et al., 2011) reported some data on IPV in the early postpartum

period (up to three months), results from each were reported differently e.g. while both Tiwari and colleagues (2005) and Zlotnick and colleagues (2011) reported scores on the CTS in the postnatal period, the latter trial reported overall scores whereas the former reported scores for separate dimensions. Other problems included trials reporting means without standard deviations and insufficient information to allow imputation of values. These are issues that apply across the gamut of outcomes and highlight the importance of comprehensive reporting of results, planning optimal follow-up schedules and familiarizing with approaches used in other similar trials.

But when is it reasonable to expect to observe a reduction in violence outcomes following intervention? For example, is one-year follow-up sufficient to capture an effect on violence? In distinguishing short- and long-term outcomes, Sullivan (1997) cautioned against hypothesizing that interventions targeting women will necessarily decrease violence in their lives. The use of IPV recurrence as a primary outcome can be problematic in the short-term on the basis that women have no control over the violence (Feder et al., 2009). Nevertheless, if a program or intervention aims to prevent further violence, proxy outcomes like safety planning are insufficient evidence of effectiveness. Another pitfall in judging a program's effect on violence is the spontaneous reduction in violence over time in a proportion of participants. Women often experience reduction in violence (and improvement in health) after they leave IPV programs even without a specific intervention (Sutherland, Bybee, & Sullivan, 1998). The same holds for women who experience IPV during pregnancy (Parker et al., 1999). McFarlane and colleagues (2006) reported that women in both arms of their trial experienced improvement on all outcomes, including threats, assaults, homicide risk and harassment at work. Finally, in relation to measuring violence outcomes, current measures may not accurately measure specific types of abuse or frequency. Evans and colleagues conducted cognitive interviews with survivors of IPV after they completed the CAS. Other than distinguishing women who have and have not experienced abuse in the follow-up period, they found variations in

interpretation and responses, which complicate its use as an outcome measure in trials. Other violence measures are likely to be equally problematic (Evans, Howarth, Gregory, Hegarty, & Feder, In Press)

4.2 Evaluation of mental health

The link between women's experiences of IPV and their mental health is well-established (Coid et al., 2006; Trevillion, Oram, Feder, & Howard, 2012). Consequently mental health has been a cornerstone of evaluation when it comes to IPV interventions in health care settings and there is an expectation that psychological interventions will deliver clinically meaningful improvements in mental health outcomes (Taft & Hegarty, 2010). When using general terms such as "depression" or "anxiety", it is important to be explicit as to whether the outcome being measured is a disorder or symptom (e.g. major depressive disorder versus depressive symptoms). Disorders are less common than symptoms, and a much larger sample is required to show a reduction in the former versus the latter. Furthermore, it is much more challenging for an intervention to have an effect on reducing incidence or prevalence of disorders versus symptoms. For these reasons, inclusion of mental health outcomes in IPV intervention research generally involves measurement of symptoms rather than disorders. However, this creates a dilemma in identifying a clinically important change in symptoms. Especially with large samples, a statistically significant change may occur, but that does not necessarily mean that the change is clinically important. It is essential that a clinically important change is specified *a priori*.

Depressive symptoms are one of the most commonly measured types of mental health problems in IPV intervention research, included as a primary or secondary outcomes across a number of the trials cited in this article (Hegarty et al., 2013; Koopman et al., 2005; Labrador et al., 2006; MacMillan et al., 2009; Tiwari et al., 2010; Tiwari et al., 2005). The Beck Depression Inventory (Beck, Steer, &

Garbin, 1988) and Center for Epidemiologic Studies-Depression Scale (Radloff, 1977), recommended for use in IPV trials (Sullivan, 1997), have been the most frequently used tools. Tiwari et al. (2005) used the Edinburgh Post-natal Depression Scale (EPDS) (Cox, Holden, & Sagovsky, 1987) with their sample of pregnant women. Taft et al. (2011) also used the EPDS as it has been validated outside the postnatal period (Thorpe, 1993). Patient Health Questionnaire (PHQ-9) is another tool, currently being used in the PATH trial to measure depression severity at one-year. Despite the number of trials that have measured depressive symptoms, Jahanfar and colleagues (2013) found that there was little information on depression (and stress) in pregnancy and the postnatal period, selected as secondary outcomes in their review of interventions in pregnancy.

As one of the most common mental health sequelae of IPV (Trevillion et al., 2012), it is not surprising that post-traumatic stress disorder (PTSD) is frequently assessed in trials. Some IPV interventions even specifically targeted PTSD (Arinero & Crespo, 2004; Edward S. Kubany et al., 2004; Labrador et al., 2006). The Clinician-administered PTSD Scale (CAPS) (Blake et al., 1990) has been used across a number of trials (Kubany et al., 2004; Resick et al., 2008), and consists of a structured interview for assessing the symptoms of PTSD according to criteria in DSM-IV. Resick and colleagues also used the Post-traumatic Diagnostic Scale (Foa, Cashman, Jaycox, & Perry, 1997) while evaluation of Kubany and colleagues' cognitive therapy-based interventions for survivors with PTSD also included the Distressing Event Questionnaire (Kubany, Leisen, Kaplan, & Kelly, 2000).

In selecting mental health measures, it is worth considering tools that combine variables, such as the Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983) used by Hegarty and colleagues (2013) and Depression Anxiety Stress Scales (Lovibond, Roznowski, & Necowitz, 1993). The Brief Symptom Inventory (BSI) (Derogatis & Spencer, 1982) has been used to assess depression (Gilbert et al., 2006) and distress (Constantino et al., 2005). Distress has also

been measured using the SCL-90-R (Hyman, 2001; Jouriles et al., 2009), which asks respondents to indicate how much they have been bothered by symptoms of fear, anxiety, depression, and somatic complaints in the past week. One intervention effect pathway is that reducing violence leads to improved mental health and wellbeing. However, interventions benefit women in various ways depending on characteristics of the participants (e.g. survivors only, women using drugs/alcohol or those with PTSD) and the intervention itself. Thus, IPV reduction does not necessarily precede improvements to mental health. An intervention aimed at enhancing perceptions of the self may also influence the breadth of women's thought-action repertoires, thereby affecting their ability to mobilize protective resources and supports (Fredrickson, 2001). In this sense, improving mental health (e.g. self-efficacy) in the short-term may be an important part of achieving violence cessation in the longer-term. Further benefits may be realized given the dose-response relationship between IPV and depression (Devries et al., 2013). Therefore, hypothesizing about processes of change helps to inform the selection of secondary as well as primary outcomes, and the length of follow-up.

4.3 Physical health

The assessment of physical health outcome was observed in a small number of trials. One trial evaluated a psychological intervention of expressive writing versus neutral writing on bodily pain (Koopman et al., 2005) based on the pain scale of the Short-Form (SF)-36 (Ware & Sherbourne, 1992). Taft et al. (2009) measured general health status, again using the SF-36; Hegarty and colleagues used the physical scale on the (briefer) SF-12. Gillum et al.'s (2009) pilot study assessed chronic pain and fatigue. It is worth considering that, despite the detrimental impact of IPV on physical wellbeing (Campbell, 2002), changes in physical health outcomes may not be discernible in the short-term and will depend on the conditions examined. Further, a tool such as the SF-36 may not be sufficiently sensitive to detect changes in the kinds of health problems that affect abused women.

4.4 Quality of life

With the exception of screening trials, quality of life is increasingly selected as either a primary or secondary outcome in trials aimed at partner violence (Ramsay et al., 2009; Taft et al., 2013). For example, it was selected as the primary outcome in Klevens and colleagues (2012) trial. It may be regarded as the most woman-centered of the outcomes, and may be responsive to changes in health care, but there is still controversy regarding how best to construct reliable measures that accurately reflect what people value. Numerous trials have adopted some variation or adaptation of the SF-12/36, which focus on physical and mental health, to measure life quality (Brierley et al., 2013; Cripe et al., 2010; Falb et al., 2014; Klevens et al., 2012; Tiwari et al., 2010; Tiwari et al., 2005). Whether health-focused tools effectively tap into what women exposed to violence consider important to their quality of life has been questioned, with another portion of trials operationalizing quality of life more broadly. The WHOQOL-Bref has been the tool of choice in most of these trials, with some including all four of the tool's dimensions (environmental; psychological; social; physical) (Hegarty et al., 2013; Krishnan, Subbiah, Chandra, & Srinivasan, 2012) and others opting to evaluate an intervention based on one subscale only (e.g. MacMillan and colleagues (2009) selected the psychological subscale). This is an outcome where there is relative consistency in the measures being adopted. Therefore, in selecting effect measures, what trialists need to decide upon is the definition of quality of life most appropriate to their intervention, what the intervention aims to improve and what constitutes a meaningful change.

4.5 Parenting and child related outcomes

A meta-analysis of abused women's expectations of health care providers identified that abused women want their children's needs to be taken into account (Feder, Hutson, Ramsay, & Taket, 2006). Another trial reported that parenting issues and children's educational needs were of greater concern than problems in their abusive relationships (Tiwari et al., 2010). The authors suggested that

the women were framing their relationship issues in the context of raising children. Meaningful outcomes might therefore include women feeling that they are parenting more effectively and that their children's health and wellbeing are improved (A. Taft & Hegarty, 2010). If interventions aim to improve outcomes for children, as well as adult survivors of IPV, they need measures of children's health, behavior and wellbeing. Feder and colleagues (2009) identified four trials of interventions aimed at children (that also included the mother) (Jouriles et al., 2001; Lieberman, Ghosh Ippen, & Van Horn, 2006; McFarlane, Groff, O'Brien, & Watson, 2005; Sullivan, Bybee, & Allen, 2002). The Child Behavior Checklist (CBCL) (Achenbach, 1991) has been used most frequently across these and other trials to assess children's behavior. It is based on mothers' reports of target children's behavior problems. Jouriles et al. (2001) used it to evaluate conduct problems and parenting skills following an intervention that taught mothers child management skills and provided them with support. In a 2009 replication, observation was replaced by the Parenting Dimensions Inventory (Power, 1993) to capture the impact of intervention on parenting and the Eyberg Child Behavior Inventory (Eyberg & Ross, 1978) was added (Ernest N. Jouriles et al., 2009). Sullivan and colleagues' (2002) trial reported on children's scores on self-competence and witnessing of IPV. Another study examined the effectiveness of child-parent psychotherapy compared to case management plus referrals based on children's PTSD symptoms and behavior problems at 6 months (Lieberman et al., 2006). A trial of a child-plus-mother intervention (Graham-Bermann, Lynch, Banyard, Devoe, & Halabu, 2007) used the CBCL and the Attitudes about Family Violence Scale (Graham-Bermann, 1994). In MOSAIC, Taft and colleagues (2011) assessed the mother-child relationship using the Parenting Stress Index Short Form-attachment subscale (Abidin, 1995).

Parenting measures can also function as intermediate outcomes (Taft & Hegarty, 2010). There are major differences between the assessment of parenting by someone other than the parent (e.g. use of an observational measure) versus the measurement of a woman's perceptions of her parenting.

Generally the latter are considered more subject to bias and less accurate. There is a wide range of tools available and child outcomes are increasingly included in trials; however, more understanding about what outcomes women want for themselves and their children from programs that include screening or other health care interventions is needed (Feder et al., 2009). For interventions to reduce/prevent IPV in antenatal settings, Jahanfar and colleagues (2013) found little information on outcomes for babies with just one (Kiely et al., 2010) out of nine included trials reporting on birthweight and preterm birth.

5. Process evaluation

Aside from collecting data on intermediate outcomes and the trial endpoints to comprehensively test hypotheses, gathering quantitative information around implementation of interventions (e.g. uptake rates) in combination with qualitative information (e.g. interviewing participants to assess barriers and facilitators to uptake and adherence of an intervention) is essential, but relatively unusual in published trials (O'Doherty et al., 2014). This type of 'process evaluation' helps to distinguish between a failure of the intervention concept or theory, and what has been referred to as an implementation failure (Oakley, Strange, Bonell, Allen, & Stephenson, 2006). Collecting data related to implementation should be relatively straightforward once planned effectively at the outset. It may involve gaining the agreement of government, clinic administrative staff or clinician participants to record certain pieces of information (e.g. numbers attending; duration of sessions), the use of surveys for participants and/or audits of medical records. Using a mixed-method approach, during and at the conclusion of the trial, is likely to generate the richest insights into processes. However, constant attention to avoiding the contamination of outcome data is warranted e.g. ensuring all final data are 'in' before approaching participants regarding process-focused interviews. Building a comprehensive process evaluation into a trial can provide insights that encourage the optimal use of future research funding.

5.1 Adverse outcomes

Harm and adverse outcomes are important aspects of any IPV trial. In addressing harm within advocacy trials, Ramsay and colleagues (2009) noted that a woman in the control group of McFarlane and colleagues' (2002) trial died by suicide 3 weeks into the trial. A woman in the intervention arm of Sullivan and colleagues (1992) was murdered by her partner. These clearly are adverse events that need to be understood in the context of the lives of the people who agree to participate in this kind of research. Data monitoring committees can facilitate discussing these rare and tragic events over the course of the trial. Though rarely reported (Ramsay et al., 2009), one approach to assessing harm is to monitor the intervention group for reduced benefit compared to the control group – e.g. increased bodily pain among women in the intervention has been observed (Koopman et al., 2005; Tiwari et al., 2005). However, the deliberate and systematic measurement of harm and adverse events in any type of IPV intervention trial is very rare, possibly due to the paucity of approaches/measures (see Valpied et al., this issue). The assessment of harm within trials needs to be extended beyond merely monitoring the differences in outcomes between groups. In the screening trial by MacMillan and colleagues (2009), the Consequences of Screening Trial (COST) instrument was developed following a comprehensive search and consultation with experts which revealed no approach to measuring harm of intervention within the violence field, or the field of health care interventions generally. The COST had undergone only preliminary testing at the time it was used in the trial by MacMillan and colleagues. A modified version was used in the WEAVE trial by Hegarty and colleagues (reported by Valpied et al. in this issue) and the MOVE trial (Taft et al., 2012). Such an approach enables assessment of the impact of participating in the research (e.g. response burden, re-traumatization, stigmatization, confidentiality breaches, impact of judgments by health providers, consequences of partner being alerted to women's involvement in trial) as well as adverse effects associated with receiving the intervention.

Recommendations and conclusions

1. In designing an evaluation framework for an IPV trial in health care settings, specify the hypothesized process by which intervention will affect the outcomes (a causal pathway). Next, identify the main outcomes (or those that are relevant to the target population, integral to clinical decision-making and can inform health policy). Select up to three as primary outcomes and others can be listed as secondary outcomes.
2. Specify (and justify) *a priori* clinically important change in primary outcome(s). This will drive your sample size.
3. Include measures for exploring the causal pathway, such as safety and readiness for change. These intermediate variables can also be included as secondary outcomes. Distinguish ~~in~~ as much as possible 'stepping stones' to the final effect(s) and potential confounding variables.
4. CONSORT guidelines (Zwarenstein et al., 2008) should inform choice of outcomes. Consider how adherence to these guidelines will affect measurement of outcomes, including approach to administration and blinding.
5. Optimize the selection of outcomes by scrutinising measures used in other trials, and determine their validity, relevance, feasibility (cost, response burden, approach administration with particular attention to safeguarding participants) in the context of IPV and with your target population. Each primary outcome, and ideally all secondary outcomes, should be measured with instruments that have established psychometric properties, including reliability and validity. Variables warranting additional psychometric development include safety, readiness for change and harm.

6. Specify (and justify) *a priori* the time frame, bearing in mind what is most appropriate to the particular outcomes and intervention at hand as well as the need for long-term outcome data. Also, ensure comprehensive reporting of results by including subscale and total scores, means and standard deviations and sufficient information to allow imputation of values.

7. Any instrument for measuring IPV needs to include physical, emotional and sexual abuse. While the measurement of IPV is essential in trials of interventions aimed at reducing IPV, with its cessation being the ultimate goal, it is important to recognize that interventions may not necessarily affect IPV in the short-term. A related matter is the need to consider the implications of instrument selection and the nature of the data generated for legal obligations to report abuse.
8. Build process evaluation into the trial from the planning phase, with particular attention to measuring harm engendered by the intervention, and by participation in the trial more generally. This includes adverse events, monitoring outcome levels in intervention versus control without introducing bias, and systematic measurement of harm.
9. This type of research presents enormous opportunities for responding to children who have been exposed to IPV and women's parenting needs. Trials need to consider developing a component of intervention for children and assessing interventions using child outcomes.
10. Consideration also needs to be given to diversity in desirable outcomes by survivors that may vary by ethnicity and sub-culture. These need to be explored through qualitative research and consultation which need to inform choice of measures.

In conclusion, there remains an urgent need for rigorous trials to determine the effectiveness of IPV interventions in low-, middle- and high-income countries. It is our view that much could be achieved in future IPV trials in respect of outcome selection and measurement to enhance the evidence base on IPV interventions in health care settings. We hope researchers will find this paper a useful resource in designing future trials, as we coordinate our efforts to tackle this major human right violation and global public health problem.

Funding: none

References

Abidin, R. R. (1995). *Parenting Stress Index: Professional Manual* (3rd ed.). Florida: PAR, Inc.

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist 4–18 and 1991 Profile.* Burlington: University of Vermont Department of Psychiatry.

Arinero, M., & Crespo, M. (2004). A treatment-outcome study of cognitive-behavioural therapy for battered women: preliminary findings. *Psicologia Conductual, 12*, 233-249.

Bair-Merritt, M. H., Jennings, J. M., Chen, R., Burrell, L., McFarlane, E., Fuddy, L., & Duggan, A. K. (2010). Reducing maternal intimate partner violence after the birth of a child: a randomized controlled trial of the Hawaii Healthy Start Home Visitation Program. *Arch Pediatr Adolesc Med, 164*(1), 16-23.

Bair-Merritt, M. H., Lewis-O'Connor, A., Goel, S., Amato, P., Ismailji, T., Jelley, M., . . . Cronholm, P. (2014). Primary care-based interventions for intimate partner violence: a systematic review. *Am J Prev Med, 46*(2), 188-194.

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8*, 77-100.

Beeble, M. L., Bybee, D., Sullivan, C. M., & Adams, A. E. (2009). Main, mediating, and moderating effects of social support on the well-being of survivors of intimate partner violence across 2 years. *J Consult Clin Psychol, 77*(4), 718-729.

Black, M. C. (2011). Intimate partner violence and adverse health consequences: implications for clinicians *American Journal of Lifestyle Medicine, 5*(5), 428-439.

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Klauminzer, G., & Charney, D. S. (1990). A clinician rating scale for assessing current and lifetime PTSD: The CAPS-1. *Behavior Therapist, 13*(187-188).

Bogat, G. A., Chin, R., Sabbath, W., & Schwartz, C. (1983). The Adult's Social Support Questionnaire (Technical Report 2). East Langsing: Michigan State University.

Bonomi, A. E., Anderson, M. L., Rivara, F. P., & Thompson, R. S. (2009). Health care utilization and costs associated with physical and nonphysical-only intimate partner violence. *Health Serv Res, 44*(3), 1052-1067.

Bradley, R., Schwartz, A.C. and Kaslow, N.J. (2005). Post-traumatic stress disorder symptoms among low-income, African American women with a history of intimate partner violence and suicidal behaviors: self-esteem, social support, and religious coping. *Journal of Traumatic Stress, 18*(6), 685-696.

Brierley, G., Agnew-Davies, R., Bailey, J., Evans, M., Fackrell, M., Ferrari, G., . . . Feder, G. S. (2013). Psychological advocacy toward healing (PATH): study protocol for a randomized controlled trial. *Trials, 14*, 221.

Bybee, D., & Sullivan, C. M. (2005). Predicting re-victimization of battered women 3 years after exiting a shelter program. *Am J Community Psychol, 36*(1-2), 85-96.

Campbell, J. C. (1986). Nursing Assessment for Risk of Homicide with Battered Women. *Advances in Nursing Science, 8*(4), 36-51.

Campbell, J. C. (1995). *Assessing Dangerousness*. Newbury Park: Sage.

Campbell, J. C. (2002). Health consequences of intimate partner violence. *Lancet, 359*, 1331-1336.

Campbell, J. C., Webster, D., & Koziol-McLain, J., et al. (2003). Risk factors for femicide in abusive relationships: results from a multisite case control study. *Am J Public Health, 93*(7), 1089-1097.

Chang, J. C., Dado, D., Ashton, S., Hawker, L., Cluss, P. A., Buranosky, R., & Scholle, S. H. (2006). Understanding behavior change for women experiencing intimate partner violence: mapping the ups and downs using the stages of change. *Patient Educ Couns, 62*(3), 330-339.

Chang, J. C., Dado, D., Hawker, L., Cluss, P. A., Buranosky, R., Slagel, L., . . . Scholle, S. H. (2010). Understanding turning points in intimate partner violence: factors and circumstances leading women victims toward change. *J Womens Health (Larchmt), 19*(2), 251-259. doi: 10.1089/jwh.2009.1568

Cluss, P. A., Chang, J. C., Hawker, L., Scholle, S. H., Dado, D., Buranosky, R., & Goldstrohm, S. (2006). The process of change for victims of intimate partner violence: support for a psychosocial readiness model. *Womens Health Issues, 16*(5), 262-274.

Cohen, S., & Williamson, G. M. (1988). *Perceived stress in a probability sample of the United States.* . Newbury Park, CA: Sage.

Coid, J., Yang, M., Roberts, A., Ullrich, S., Moran, P., Bebbington, P., . . . Singleton, N. (2006). Violence and psychiatric morbidity in a national household population--a report from the British Household Survey. *Am J Epidemiol, 164*(12), 1199-1208.

Coker, A. L., Davis, K. E., Arias, I., Desai, S., Sanderson, M., Brandt, H. M., & Smith, P. H. (2002). Physical and mental health effects of intimate partner violence for men and women. *Am J Prev Med, 23*(4), 260-268.

Constantino, R., Kim, Y., & Crane, P. A. (2005). Effects of a social support intervention on health outcomes in residents of a domestic violence shelter: a pilot study. *Issues Ment Health Nurs, 26*(6), 575-590.

Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry, 150*, 782-786.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *Bmj, 337*, a1655.

Cripe, S. M., Sanchez, S. E., Sanchez, E., Ayala Quintanilla, B., Hernandez Alarcon, C., Gelaye, B., & Williams, M. A. (2010). Intimate partner violence during pregnancy: a pilot intervention program in Lima, Peru. *J Interpers Violence, 25*(11), 2054-2076. doi: 10.1177/0886260509354517

Curry, M. A., Durham, L., Bullock, L., Bloom, T., & Davis, J. (2006). Nurse case management for pregnant women experiencing or at risk for abuse. *J Obstet Gynecol Neonatal Nurs, 35*(2), 181-192.

Derogatis, L. R., & Spencer, P. (1982). The brief symptom inventory (BSI) administration, scoring and procedures manual-I. *Clinical Psychometric Research*.

Devries, K., Mak, J., Bacchus, L., Child, J., Falder, G., Petzold, M., . . . Watts, C. (2013). Intimate partner violence and incident depressive symptoms and suicide attempts: a systematic review of longitudinal studies. *PLoS Med, 10*(5), e1001439.

Dienemann, J., Campbell, J., Landenburger, K., & Curry, M. A. (2002). The Domestic Violence Survivor Assessment: a tool for counseling women in intimate partner violence relationships. *Patient Educ Couns, 46*(3), 221-228.

Dutton, M. (1992). *Empowering and healing the battered woman*. New York: Springer.

Evans, M., Howarth, E., Gregory, A., Hegarty, K., & Feder, G. (In Press). Even ‘daily’ is not enough”: How well do we measure domestic violence and abuse? A thinkaloud study of a commonly used self-report scale.

Eyberg, S. M., & Ross, A. W. (1978). Assessment of child behavior problems: The validation of a new inventory. *Journal of Clinical Child Psychology, 7*, 113-116.

Falb, K. L., Diaz-Olavarrieta, C., Campos, P. A., Valades, J., Cardenas, R., Carino, G., & Gupta, J. (2014). Evaluating a health care provider delivered intervention to reduce intimate partner violence and mitigate associated health risks: study protocol for a randomized controlled trial in Mexico City. *BMC Public Health, 14*, 772. doi: 10.1186/1471-2458-14-772

Feder, G., Davies, R., Baird, K., Dunne, D., Eldridge, S., Griffiths, C., . . . Sharp, D. (2011). Identification and Referral to Improve Safety (IRIS) of women experiencing domestic violence with a primary care training and support programme: a cluster randomised controlled trial. *Lancet, 378*(9805), 1788-1795.

Feder, G., Hutson, M., Ramsay, J., & Taket, A. (2006). Women exposed to intimate partner violence: expectations and experiences when they encounter health care professionals: a meta-analysis of qualitative studies. *Archives of Internal Medicine, 166*(1), 22-37.

Feder, G., Ramsay, J., Dunne, D., Rose, M., Arsene, C., Norman, R., . . . Taket, A. (2009). How far does screening women for domestic (partner) violence in different health-care settings meet criteria for a screening programme? Systematic reviews of nine UK National Screening Committee criteria. *Health Technology Assessment, 13*(16), iii-iv, xi-xiii, 1-113, 137-347.

Foa, E. B., Cashman, L. A., Jaycox, L., & Perry, K. (1997). The validation of a self-report measure of posttraumatic stress disorder: The Posttraumatic Diagnostic Scale. *Psychological Assessment, 4*, 445-451.

Ford-Gilboe, M., Merritt-Gray, M., Varcoe, C., & Wuest, J. (2011). A theory-based primary health care intervention for women who have left abusive partners. *ANS Adv Nurs Sci, 34*(3), 198-214.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology. The broaden-and-build theory of positive emotions. *Am Psychol, 56*(3), 218-226.

Fry, P. S., & Barker, L. A. (2002). Female Survivors of abuse and violence: The influence of Storytelling Reminiscence on Perceptions of Self-efficacy, Ego Strength, and Self-esteem. In J. D. Webster & B. K. Haight (Eds.), *Critical Advances in Reminiscence Work* (pp. 197-217): Springer Publishing Company.

Gilbert, L., El-Bassel, N., Manuel, J., Wu, E., Go, H., Golder, S., . . . Sanders, G. (2006). An integrated relapse prevention and relationship safety intervention for women on methadone: testing short-term effects on intimate partner violence and substance use. *Violence Vict, 21*(5), 657-672.

Gillum, T. L., Sun, C. J., & Woods, A. B. (2009). Can a Health Clinic-Based Intervention Increase Safety in Abused Women? Results from a Pilot Study. *Journal of Womens Health, 18*(8), 1259-1264. doi: 10.1089/jwh.2008.1099

Graham-Bermann, S. A. (1994). *Attitudes About Family Violence (AAFV) Scale*. Department of Psychology, University of Michigan.

Graham-Bermann, S. A., Lynch, S., Banyard, V., Devoe, E. R., & Halabu, H. (2007). Community-based intervention for children exposed to intimate partner violence: An efficacy trial. *Journal of Consulting and Clinical Psychology, 75*(2), 199-209. doi: 10.1037/0022-006x.75.2.199

Gunasekara, F. I., Carter, K., & Blakely, T. (2008). Glossary for econometrics and epidemiology. *J Epidemiol Community Health, 62*(10), 858-861.

Hegarty, K., Bush, R., & Sheehan, M. (2005). The Composite Abuse Scale: further development and assessment of reliability and validity of a multidimensional partner abuse measure in clinical settings. *Violence Vict, 20*(5), 529-547.

Hegarty, K., O'Doherty, L., Gunn, J., Pierce, D., & Taft, A. (2008). A brief counseling intervention by health professionals utilising the 'readiness to change' concept for women experiencing intimate partner abuse: The weave project. *J Fam Stud, 14*(2-3), 376-388.

Hegarty, K., O'Doherty, L., Taft, A., Chondros, P., Brown, S., Valpied, J., . . . Gunn, J. (2013). Screening and counselling in the primary care setting for women who have experienced intimate partner violence (WEAVE): a cluster randomised controlled trial. *Lancet, 382*(9888), 249-258.

Heise, L., Ellsberg, M., & Gottmoeller, M. (2002). A global overview of gender-based violence. *Int J Gynaecol Obstet, 78 Suppl 1*, S5-14.

Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [Updated March 2011]*: The Cochrane Collaboration.

Hudson, W. (1982). *The clinical measurement package*. Homewood, IL: Dorsey.

Hyman, K. (2001). *Impact of emergency department intimate partner violence (IPV) advocacy: A longitudinal, randomized trial [PhD thesis]*. University of Pittsburgh. Pittsburgh, PA.

Jahanfar, S., Janssen, P. A., Howard, L. M., & Dowswell, T. (2013). Interventions for preventing or reducing domestic violence against pregnant women. *Cochrane Database Syst Rev, 2*, CD009414.

Jouriles, E. N., McDonald, R., Rosenfield, D., Stephens, N., Corbitt-Shindler, D., & Miller, P. C. (2009). Reducing conduct problems among children exposed to intimate partner violence: a randomized clinical trial examining effects of Project Support. *J Consult Clin Psychol, 77*(4), 705-717.

Jouriles, E. N., McDonald, R., Spiller, L., Norwood, W. D., Swank, P. R., Stephens, N., . . . Buzy, W. M. (2001). Reducing conduct problems among children of battered women. *J Consult Clin Psychol, 69*(5), 774-785.

Kiely, M., El-Mohandes, A. A. E., El-Khorazaty, M. N., & Gantz, M. G. (2010). An integrated intervention to reduce intimate partner violence in pregnancy: a randomized controlled trial. *Obstet Gynecol, 115*(2 Pt 1), 273-283.

Kirkwood, C. (1993). *Leaving abusive partners*. London: Sage.

Klevens, J., Kee, R., Trick, W., Garcia, D., Angulo, F. R., Jones, R., & Sadowski, L. S. (2012). Effect of screening for partner violence on women's quality of life: a randomized controlled trial. *Jama, 308*(7), 681-689. doi: 1346180 [pii]

10.1001/jama.2012.6434

Koopman, C., Ismailji, T., Holmes, D., Classen, C. C., Palesh, O., & Wales, T. (2005). The effects of expressive writing on pain, depression and posttraumatic stress disorder symptoms in survivors of intimate partner violence. *J Health Psychol, 10*(2), 211-221.

Koziol-McLain, J., Garrett, N., Fanslow, J., Hassall, I., Dobbs, T., Henare-Toka, T. A., & Lovell, V. (2010). A randomized controlled trial of a brief emergency department intimate partner violence screening intervention. *Annals of Emergency Medicine, 56*(4), 413-423 e411. doi: S0196-0644(10)00450-6 [pii]

10.1016/j.annemergmed.2010.05.001

Krishnan, S., Subbiah, K., Chandra, P., & Srinivasan, K. (2012). Minimizing risks and monitoring safety of an antenatal care intervention to mitigate domestic violence among young Indian women: The Dil Mil trial. *BMC Public Health, 12*, 943. doi: 10.1186/1471-2458-12-943

Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *Lancet, 360*(9339), 1083-1088.

Kubany, E. S., Hill, E. E., & Owens, J. A. (2003). Cognitive trauma therapy for battered women with PTSD: preliminary findings. *J Trauma Stress, 16*(1), 81-91.

Kubany, E. S., Hill, E. E., Owens, J. A., Iannce-Spencer, C., McCaig, M. A., Tremayne, K. J., & Williams, P. L. (2004). Cognitive trauma therapy for battered women with PTSD (CTT-BW). *J Consult Clin Psychol, 72*(1), 3-18.

Kubany, E. S., Leisen, M. B., Kaplan, A. S., & Kelly, M. P. (2000). Validation of a brief measure of posttraumatic stress disorder: the Distressing Event Questionnaire (DEQ). *Psychol Assess, 12*(2), 197-209.

Labrador, F. J., Fernandez-Velasco, M. D., & Rincon, P. P. (2006). Efficacy of a brief and individual participation program by posttraumatic stress disorder in female victims of domestic violence. *Int J Clin Health Psychol, 6*, 527-547.

Liang, B., Goodman, L., Tummala-Narra, P., & Weintraub, S. (2005). A theoretical framework for understanding help-seeking processes among survivors of intimate partner violence. *Am J Community Psychol, 36*(1-2), 71-84.

Lieberman, A. F., Ghosh Ippen, C., & Van Horn, P. (2006). Child-parent psychotherapy: 6-month follow-up of a randomized controlled trial. *J Am Acad Child Adolesc Psychiatry, 45*(8), 913-918.

Lovibond, S. H., Roznowski, M., & Necowitz, L. B. (1993). Manual for the Depression Anxiety Stress Scales *Psychology Foundation Monograph*: University of New South Wales, NSW.
MacMillan, H. L., Fleming, J. E., Streiner, D. L., Lin, E., Boyle, M. H., Jamieson, E., . . . Beardslee, W. R. (2001). Childhood abuse and lifetime psychopathology in a community sample. *Am J Psychiatry, 158*(11), 1878-1883.
MacMillan, H. L., Wathen, C. N., Jamieson, E., Boyle, M., McNutt, L. A., Worster, A., . . . Webb, M. (2006). Approaches to screening for intimate partner violence in health care settings: a randomized trial. *Jama, 296*(5), 530-536.
MacMillan, H. L., Wathen, C. N., Jamieson, E., Boyle, M. H., Shannon, H. S., Ford-Gilboe, M., . . . McNutt, L. A. (2009). Screening for intimate partner violence in health care settings: a randomized trial. *Jama, 302*(5), 493-501.
Mancoske, R. J., Standifer, D., & Cauley, C. (1994). The effectiveness of brief counseling services for battered women. *Res Social Work Pract, 4*(1), 53-63.
McFarlane, Groff, J. Y., O'Brien, J. A., & Watson, K. (2006). Secondary prevention of intimate partner violence: a randomized controlled trial. *Nurs Res, 55*(1), 52-61.
McFarlane, Parker, B., Soeken, K., Silva, C., & Reel, S. (1998). Safety behaviors of abused women after an intervention during pregnancy. *J Obstet Gynecol Neonatal Nurs, 27*(1), 64-69.
McFarlane, Soeken, K., & Wiist, W. (2000). An evaluation of interventions to decrease intimate partner violence to pregnant women. *Public Health Nurs, 17*(6), 443-451.
McFarlane, J. M., Groff, J. Y., O'Brien, J. A., & Watson, K. (2005). Behaviors of children exposed to intimate partner violence before and 1 year after a treatment program for their mother. *Appl Nurs Res, 18*(1), 7-12.
Nelson, H. D., Bougatsos, C., & Blazina, I. (2012). Screening women for intimate partner violence: A systematic review to update the 2004 Preventive Services Task Force recommendation. *Ann Intern Med, 156*(11), 1-13; W11-W13.
O'Doherty, L. J., Taft, A., Hegarty, K., Ramsay, J., Davidson, L. L., & Feder, G. (2014). Screening women for intimate partner violence in healthcare settings: abridged Cochrane systematic review and meta-analysis. *Bmj, 348*, g2913.
Oakley, A., Strange, V., Bonell, C., Allen, E., & Stephenson, J. (2006). Process evaluation in randomised controlled trials of complex interventions. *Bmj, 332*(7538), 413-416. doi: 332/7538/413 [pii]
10.1136/bmj.332.7538.413
Parker, B., McFarlane, J., Soeken, K., Silva, C., & Reel, S. (1999). Testing an Intervention To Prevent Further Abuse to Pregnant Women. *Research in Nursing and Health, 22*, 59-66.
Power, T. G. (1993). *Parenting Dimensions Inventory (PDI): A research manual*. Houston, Tx: Psychology Department, University of Houston.
Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas., 1*, 385-401.
Ramsay, J., Carter, Y., Davidson, L., Dunne, D., Eldridge, S., Feder, G., . . . Warburton, A. (2009). Advocacy interventions to reduce or eliminate violence and promote the physical and psychosocial well-being of women who experience intimate partner abuse (Review). *Cochrane Database of Systematic Reviews*(3). doi: Cd005043
10.1002/14651858.CD005043.pub2
Reisenhofer, S., & Taft, A. (2013). Women's journey to safety - the Transtheoretical model in clinical practice when working with women experiencing intimate partner violence: a scientific review and clinical guidance. *Patient Educ Couns, 93*(3), 536-548.
Resick, P. A., Galovski, T. E., Uhlmansiek, M. O., Scher, C. D., Clum, G. A., & Young-Xu, Y. (2008). A Randomized clinical trial to dismantle components of cognitive processing therapy for posttraumatic stress disorder in female victims of interpersonal violence. *Journal of Consulting and Clinical Psychology, 76*(2), 243-258. doi: 10.1037/0022-006x.76.2.243

Rosenberg, M. (1965). Rosenberg Self-Esteem Scale, The Morris Rosenberg Foundation.
Scherer, M. (1982). The self-efficacy scale: construction and validation. *Psychological Reports, 51*, 663-671.
Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35-37). Windsor, UK: NFER-NELSON.
Seuc, A. H., Peregoudov, A., Betran, A. P., & Gulmezoglu, A. M. (2013). Intermediate outcomes in randomized clinical trials: an introduction. *Trials, 14*, 78.
Sullivan, C. M. (1997). Victim safety and well being: measure of short-term and long-term change. In M. Burt, A. Harrell, L. Newmark, L. Aron & L. Jacobs (Eds.), *The Urban Institute Evaluation Guidebook: Projects funded by STOP Formula Grants Under the Violence Against Women*: National Institute of Justice.
Sullivan, C. M., Bybee, D. I., & Allen, N. E. (2002). Findings from a community-based program for battered women and their children. *J Interpers Violence, 17*(9), 915-936.
Sullivan, C. M., Tan, C., Basta, J., Rumptz, M., & Davidson, W. S., 2nd. (1992). An advocacy intervention program for women with abusive partners: initial evaluation. *Am J Community Psychol, 20*(3), 309-332.
Sutherland, C., Bybee, D., & Sullivan, C. (1998). The long-term effects of battering on women's health. *Womens Health, 4*(1), 41-70.
Suvak, M. K., Taft, C. T., Goodman, L. A., & Dutton, M. A. (2013). Dimensions of functional social support and depressive symptoms: a longitudinal investigation of women seeking help for intimate partner violence. *Journal of Consulting and Clinical Psychology, 81*(3), 455-466.
Taft, O'Doherty, L., Hegarty, K., Ramsay, J., Davidson, L., & Feder, G. (2013). Screening women for intimate partner violence in healthcare settings. *Cochrane Database Systematic Reviews, 4*, CD007007.
Taft, A., & Hegarty, K. (2010). Intimate partner violence against women: What outcomes are meaningful? *Jama, 304*(5), 577-579.
Taft, A. J., Small, R., Hegarty, K. L., Watson, L. F., Gold, L., & Lumley, J. (2011). Mothers' AdvocateS In the Community (MOSAIC) - non-professional mentor support to reduce intimate partner violence and depression in mothers: a cluster randomised trial in primary care *BMC Public Health, 11*(178), 1-10.
Taft, A. J., Small, R., Humphreys, C., Hegarty, K., Walter, R., Adams, C., & Agius, P. (2012). Enhanced maternal and child health nurse care for women experiencing intimate partner/family violence: protocol for MOVE, a cluster randomised trial of screening and referral in primary health care. *BMC Public Health, 12*, 811.
Thompson, M., Basile, K., Hertz, M., & Sitterle, D. (2006). *Measuring intimate partner violence victimisation and perpetration. A compendium of assessment tools*. Atlanta, GA: Centers for Disease Controla nd Prevention, National Center for Injury Preventions and Control.
Thorpe, K. (1993). A study of the use of the Edinburgh postnatal depression scale with parent groups outside the postpartum period. *Journal of Reproductive and Infant Psychology, 11*, 119-125.
Tiwari, A., Fong, D. Y., Wong, J. Y., Yuen, K. H., Yuk, H., Pang, P., . . . Bullock, L. (2012). Safety-promoting behaviors of community-dwelling abused Chinese women after an advocacy intervention: a randomized controlled trial. *Int J Nurs Stud, 49*(6), 645-655.
Tiwari, A., Fong, D. Y., Yuen, K. H., Yuk, H., Pang, P., Humphreys, J., & Bullock, L. (2010). Effect of an advocacy intervention on mental health in Chinese women survivors of intimate partner violence: a randomized controlled trial. *Jama, 304*(5), 536-543.
Tiwari, A., Leung, W. C., Leung, T. W., Humphreys, J., Parker, B., & Ho, P. C. (2005). A randomised controlled trial of empowerment training for Chinese abused pregnant women in Hong Kong. *Bjog, 112*(9), 1249-1256.

Tiwari, A., Yuk, H., Pang, P., Fong, D. Y., Yuen, F., Humphreys, J., & Bullock, L. (2012). Telephone intervention to improve the mental health of community-dwelling women abused by their intimate partners: a randomised controlled trial. *Hong Kong Med J, 18 Suppl 6*, 14-17.

Trevillion, K., Oram, S., Feder, G., & Howard, L. (2012). Experiences of domestic violence and mental disorders: a systematic review and meta-analysis. *PLoS ONE, 7*(12), e51740.

Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care, 30*(6), 473-483.

Wathen, C. N., Jamieson, E., & MacMillan, H. L. (2008). Who is identified by screening for intimate partner violence? *Womens Health Issues, 18*(6), 423-432.

Wathen, C. N., & MacMillan, H. L. (2012). Health care's response to women exposed to partner violence: moving beyond universal screening. *Jama, 308*(7), 712-713. doi: 1346157 [pii]
10.1001/jama.2012.9913

World Health Organization. (2013a). Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence. Geneva: WHO

World Health Organization. (2013b). Responding to intimate partner violence and sexual violence against women: WHO clinical and policy guidelines. Geneva: WHO.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatr Scand, 67*(6), 361-370.

Zink, T., Elder, N., Jacobson, J., & Klostermann, B. (2004). Medical management of intimate partner violence considering the stages of change: Precontemplation and contemplation. *Ann Fam Med, 2*(3), 231-239.

Zlotnick, C., Capezza, N. M., & Parker, D. (2011). An interpersonally based intervention for low-income pregnant women with intimate partner violence: a pilot study. *Arch Womens Ment Health, 14*(1), 55-65.

Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D. G., Tunis, S., Haynes, B., . . . Moher, D. (2008). Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *Bmj, 337*, a2390.