

# Evaluating contracted domestic violence programs

## Standardisation and organisational culture<sup>1</sup>

The growing significance of domestic violence programs run by the state and contracted non-government agencies in Australia over the past decade has not been matched by a commensurate growth in evaluation of those programs. A number of evaluations have been conducted, but only a few have been long-term and large-scale, due to funding and other constraints. The promotion of evidence-based practice and policy encourages practitioners and scholars to aim for comprehensive and systematic review of initiatives, but achieving that is no mean feat since both the service delivery and governance of these programs are often complex and multi-stranded.

In this article we reflect on theoretical and practical issues of evaluation by reporting on the experience of undertaking an evaluation of domestic violence perpetrator and victim support programs delivered for an Australian state government by contracted community service agencies. The experience was not as we had expected on the basis of social policy debates. In this situation, there were difficulties in generating sufficiently robust data on offenders in group programs that delayed and limited the scope of a quasi-experimental assessment of intervention, although that strand continues to date. At the same time, however, this strengthened the importance of evaluating service delivery and governance issues within a 'realist evaluation' paradigm. Since contemporary domestic violence programs often typically combine coordinated support for women and children and behaviour change intervention with offenders, evaluation needs multi-stranded approaches that may draw on paradigms more commonly thought of as being in tension.

### Introduction

Private violence has become a public issue over the past few decades and in Australia. Two of the more tangible manifestations of this shift have been the development of emergency accommodation services for women and children and domestic violence legislation. Recent funding initiatives include the joint

**Ed Carson**

**Donna Chung**

**Andrew Day**



*Ed Carson (top left) is Professor of Social Policy in the School of Psychology, Social Work and Social Policy at the University of South Australia, Magill Campus, Adelaide. Email: <edgar.carson@unisa.edu.au>*

*Donna Chung (top right) is an Associate Professor in Social Work and Director of the Centre for the Study of Safety and Wellbeing at the University of Warwick, Coventry, England. Email: <donna.chung@warwick.ac.uk >*

*Andrew Day (bottom) is an Associate Professor at the Centre for Offender Reintegration at Deakin in the School of Psychology, Faculty of Health, Medicine, Nursing and Behavioural Sciences at Deakin University, Geelong. Email: <andrew.day@deakin.edu.au>*

Commonwealth, states and territories' *Partnerships Against Domestic Violence* (PADV) program. This has been responsible for funding a raft of projects to improve practice, and share knowledge about what is deemed to be good practice. A range of these initiatives included research and evaluation projects. However, it is debatable whether the evidence in the field is robust and, moreover, the way good practice is determined remains somewhat contentious.

This article reports on the experience of a team of university researchers undertaking an evaluation of a small number of domestic violence perpetrator and victim support programs delivered by not-for-profit community service agencies contracted to do so by an Australian state government. The aim of the project was to compare different models of intervention across urban and rural sites, review the methods of intervention that were used to promote the safety of women and children victims, and evaluate the outcomes of the programs. At this stage of the analysis, only preliminary observations are available on the effectiveness of the programs for either perpetrators or victims, but the research has raised a number of more general methodological issues about evaluation of domestic violence programs.

In this article we initially outline some of the key debates about common evaluation paradigms in the human services and then discuss ethical and practical issues that have a bearing on these differing conceptual and disciplinary perspectives. Given the increasing trends towards government departments contracting third-sector agencies to deliver domestic violence services (as part of a broader trend of contracting out welfare services), it is necessary to discuss how funding and reporting requirements associated with contractualism might have a bearing on program design, data collection and assessment of programs. This requires consideration of whether tighter reporting regimes associated with current funding of domestic violence programs generate particular forms of output and outcome data that are more appropriate for some evaluation approaches than others.

These findings and trends are then considered in the light of the authors' current evaluation of domestic violence perpetrator programs that are contracted to third-sector agencies by the state. The article contrasts the literature with our experiences of evaluating the programs, in order to suggest some implications for the current, and future, domestic violence program evaluation.

## Evaluation and evidence

### Macro and meta-evaluation

Evidence-based policy and practice are the touchstones of recent debates about social welfare provision, notwithstanding heated debates about what constitutes evidence, how one prioritises different types of data and what methodological paradigms are privileged in these rankings. This is evident in the dialogue between promoters of the Cochrane and Campbell Collaborations (Campbell

Collaboration 2009) and their critics, as discussed in Pawson (2006) and, more generally, in Meagher (2002).

Assessment of the social impact of domestic violence programs is theoretically and administratively possible at the macro level by documenting correlations between a program introduction and changes in crime rates, drug use, recidivism etc., and at the meso-level by reporting on consequences of participation for particular categories of participants. Macro-level evaluation emphasising victim safety is also possible using, for example, measures of unmet demand for services. The Australian Institute of Health and Welfare, for example, publishes data on unmet demand for domestic violence shelters as a subset of reporting on SAAP programs, but constraints on reporting make this a vexed area of assessment (AIHW 2008). It is argued by critics that such high level or abstracted measures do not accurately reflect effectiveness of domestic violence responses because they overlook some key factors about domestic violence (Pence & Shepard 1999). The majority of domestic violence cases are not reported to police; criminal charges related to domestic violence are not always pursued successfully to prosecution and many women do not seek emergency accommodation in escaping domestic violence. Therefore, critics argue that there is a need for local level data and perspectives in evaluating the impact of domestic violence responses, as the context or environment is important (Klevens et al. 2008). Nevertheless, large-scale quantitative methodologies are available to assess program take-up and benefits, including cost-benefit analysis, and are used occasionally in domestic violence research.

In addition, meso-level evaluation approaches are possible by reporting the impact of a perpetrator behaviour change program on a pool of offenders, by measuring changes in the flow of conviction/referral figures over the duration of the program, as has been undertaken by researchers such as Edward Gondolf in North America (Gondolf 2002) and others, for example Hendricks et al. (2006). Such approaches are complex to administer even where the participation rates are high and outcomes are easy to identify (notwithstanding debates in the specialist literature about definitions and measurement of outcomes). In domestic violence contexts these preconditions for robust measurement often do not exist.

In a similar vein, PADV commissioned an economic costing of domestic violence (Access Economics 2004) and this can be useful in further judging whether a certain level of expenditure is producing relatively more changes on a pool of offenders than expenditure at another level. Even though sizeable budgets have been allocated to them in countries such as Australia, such flow analysis is difficult to apply to violence prevention programs that have relatively few numbers of participants.

Shifting focus to the recording and classification of outcomes of programs is even more difficult,

especially where programs are particularly concerned with victim safety and they entail strategies tailored to victim needs that are relatively variable and open-ended. Nonetheless, some evaluations of domestic violence intervention in Australia have been concerned with such an analysis of procedures at the program level, rather than macro-evaluation of domestic violence initiatives (Szirom, Chung & Jaffe 2003).

While the current systems are relatively complex, program- and project-level comparisons across services and programs are technically feasible, even though they need to be handled with caution. Meso- and micro-evaluations can provide basic information on program processes, such as the number of participants who enter and leave a program over a certain period. Even when such data are entirely descriptive, this can be useful to determine whether the involvement of target beneficiaries has reached or fallen short of the goals set. It can also provide information about aspects of strengths and weaknesses of the agencies and overall system processes.

When the focus switches to outcomes of programs rather than just usage or throughput, however, it is harder because comprehensive evaluations seek to go beyond descriptive data, to establish the counterfactual (Cummings 2006). That is, they need to determine whether desired changes did happen, but also what would have happened in the absence of the program, and whether the assistance for one group is at the expense of another group. Typically, the net impact of programs aiming to produce social and behavioural change is smaller than the proportion of participants who report changed attitudes and behaviour after the program due to issues of 'deadweight' (sometimes called the 'windfall' factor) where an outcome would have occurred even in the absence of the program. Thus the key concern for program evaluation is the difference produced from participation/intervention.

A critical question is how to deal with contextual variables that influence consequences of participation in particular situations. One can either try to control for these to eliminate them, or try to acknowledge the extent to which they are critical in understanding the outcome and thus to highlight them. The article takes these emphases in turn.

#### **Establishing the counter-factual: experimental and quasi-experimental design**

For domestic violence perpetrator programs, where outcomes are potentially easier to measure than larger scale interventions, an experimental design is the approach of choice for small-scale evaluations that assess attitudinal and behavioural change from a psychological perspective. Controlling for external variables could be achieved by random assignment or explicit matching of a control group of non-participants with the same characteristics as the participant ('experimental') group—age, gender, education are typical comparison criteria, although

one cannot be confident that differences observed between two groups after a program are attributable to the program rather than to pre-existing differences between participants' and non-participants' unmeasured characteristics (Trochim 2000).

In any case, in Australia it is not at all common to establish control groups in evaluations of offender programs because there may be strong ethical and legal reasons why eligible applicants should not be denied access or relieved of the obligation to attend offender rehabilitation programs. While it is technically possible, the use of the experimental method with random assignment is confined either to projects that are declared to be pilot projects or demonstration projects that, it is argued, are not depriving citizens of rights to access programs of social assistance to which they are entitled.

Consequently, a quasi-experimental methodology is more likely to be used to analyse programs by comparing the experiences of different categories of program participants. This is a common, indeed the most common, approach to evaluation of many social programs in Australia, as in most OECD countries (Shadish, Cook & Campbell 2002). This is acknowledged by researchers who are participants in the Campbell Collaboration, the international research network that produces systematic reviews of the effects of social interventions (Campbell Collaboration 2009). They are, by definition, strong advocates of evidence-based practice in social policy, based on systematic reviews and an associated research hierarchy that privileges experimental research design. They observe, for example, '... there is something compelling about these [quasi-experimental] designs; taken as a group, they are easily more frequently implemented than their randomized cousins' (Trochim 2000). Quasi-experiments constitute natural experiments' in which individual differences between participants can be assessed and recorded over the duration of the program. Here again, the literature catalogues reasons for caution in program evaluations to date (Bowen, Bowen & Ware 2002). Sample selection/inclusion and attrition are problematic, all the more so in the domestic violence arena because it is widely accepted that those who complete treatment are less likely to re-assault than those who drop out (Gondolf 2002).

A limiting factor is that systematic reviews and meta-evaluations, which aggregate the findings from individual evaluations to find a net effect, may minimise how programs, practices and contexts succeed in their own terms because programs that produce positive consequences are lumped in with those that produce negative consequences. Where evaluation of outcomes at domestic violence program level has been undertaken in Australia, notwithstanding the debates about outcome measures used, the effectiveness of programs has been varied (O'Leary, Chung & Zannettino 2004). Even where there are observed patterns that do not cancel each other out, correlation is not causation, and a persuasive argument always needs to be

generated to account for any observed change in societal patterns. Thus much more research and evaluation needs to be done, but this re-opens the question of an appropriate methodology.

### Realist evaluation

Proponents of ‘realist evaluation’ (initially called realistic evaluation) are doubtful about the positivist methodology as a method of finding out which programs do and which do not succeed (Pawson 2006; Pawson & Tilley 1997). They argue that the classic experimental method of the natural sciences does conceptual violence to what is normal practice in social policy and social science research. In particular, in the field of domestic violence it has been argued that practice is path dependent and thus historically constrained, with the consequence that it is not feasible to undertake standardised evaluations in a positivist methodological mode.

For example, a much-discussed intervention in the domestic violence arena concerned the merits of mandatory arrest as a means of reducing rates of repeated assault (Sherman 1992, Sherman et al. 1992; Sherman, Schmidt & Rogan 1992). In the US state of Minnesota in the 1980s, three groups of police officers responding to domestic violence reports where there was no serious injury were required to arrest the perpetrator, provide advice or send the perpetrator away. There was a statistically significant lower rate of repeat calls for domestic violence among the group where arrest occurred, compared to the other groups. A number of cities in other states in the US subsequently adopted the strategy and, in some cases, the mandatory arrest policy reduced repeat domestic violence call-outs but it increased in others. Sherman (1992) suggested that this could be explained by the different community, employment and family structures in the various cities. In stable communities with high rates of employment, arrest produces shame for the perpetrator who is then less likely to re-offend—but in less stable communities it is more liable to trigger anger than shame. Thus the effect of arrest varies by context. A further consideration is that repeat call-outs could also be a positive effect in some instances, as it could also indicate that because women received a positive response from the police call-out they are more likely to call again when unsafe. While this does not constitute an indication of success in stopping the domestic violence altogether, it has offered a safe response that increased women’s safety.

A more immediate argument for contextually informed evaluation lies in the history of Australian domestic violence programs. The typical programs delivered by contracted community service agencies in Australia have not usually been generated as fully integrated services with a consistent theoretical approach. Over the last 20 years or so, a multitude of domestic violence projects have been implemented that have described themselves, variously, as ‘integrated responses’, ‘collaborative efforts’, ‘partnerships’, and so on, but few of

these programs have been newly created (O’Leary, Chung & Zannettino 2004). Rather, they typically involve the coordination of a number of agencies already involved in responses to domestic violence, sometimes from a variety of philosophical or therapeutic perspectives. Thus comprehensive evaluation of even relatively modest social change programs can be very complex because of the range of administrative, delivery and outcome issues that ideally should be considered.

Domestic violence services delivered by a number of contracted agencies in the same Australian city or state may not have the same practice frameworks and guiding interventions (Blay & Wright 2008). The agencies themselves may not have explicitly articulated their theoretical frameworks, allowing service delivery to be shaped by incremental modifications over time to what may or may not have initially been conceptually coherent strategies. Where agencies are operating in different geographical areas the pressures on service delivery integrity are likely to be quite different, making it more difficult to compare and contrast programs delivered by different agencies in different geographic areas. This is relevant for both domestic violence perpetrator strands and victim support strands of programs.

Yet current Australian funding and reporting practices are believed to shape service delivery in ways that have implications for this variability, as we discuss below.

### Pressures towards standardised tracking, recording and reporting data: contractualism

Some Australian literature on outsourcing of human services has highlighted a trend towards standardisation of services, whereby a funding department increasingly specifies the accountability of contracted agencies to produce declared outcomes, which are to be reported through standard protocols and templates (Carson & Wadham 2001; Meagher 2002). Funder expectations of services being delivered by contracted agencies are specified even more tightly in funding submission protocols and reporting templates. A key expectation of evaluators is that these trends increase the generation of standardised process and outcomes data that improve the prospects of this form of evaluation. It is argued by advocates that the increased use of contracts has the twin benefits of increasing efficiency and transparency of services delivered and also of facilitating the closely related search for evidence-based policy and practice. Some policy analysts believe this will generate practice that is more amenable to experimental and quasi-experimental evaluation, and strengthen the role of evaluation as a condition of future funding rounds.

Conversely, there is a strong strand of criticism in the social policy literature that increased outsourcing and contractualism in community

services is transforming service delivery and organisational culture of community service agencies via the funder's specification of agency deliverables and reporting obligations. The argument has been made that the trends to standardisation and conformity essentially constitute the fetishism of easily measurable outcomes, and that such funding and reporting drivers constrain the range of service options, including the political imperatives of the agencies (Egan & Hoatson 1999). Moreover, some of this literature adopts a stance that such reporting requirements may unnecessarily constrain the professionalism of workers in community service agencies (Carson, King & Mayer 2007; Meagher 2002).

Our initial experiences of the present domestic violence program evaluations do not readily conform to such expectations, however, and this warrants a more detailed consideration of the debates in the literature in order to reflect further on this situation.

The trends in funding mechanisms for welfare services in Australia over recent years have been well documented. Australia-wide, non-government community service agencies increased and expanded through contracting-out during the 1980s and 1990s and moves to adopt a 'New Public Management' model over this period are widely recognised as aiming to decouple policy from service delivery by outsourcing human service provision.<sup>2</sup> Funding regimes based on a competitive tendering model have been accompanied by an emphasis on efficiency and productivity, and increasing regulation of outputs through accountability/efficiency/governance requirements. The trend has progressed apace, with advocates arguing that contractualism gives funding departments the ability to access specialist services that:

- extend departmental capacity
- improve service quality
- increase efficiency
- offer flexibility in service delivery.

At the same time, there has been considerable discussion about constraints and challenges for both departments and agencies in such a system, including increased transaction costs for contract tendering and subsequent accountability and reporting requirements. More specifically, during that period of growth it has been argued that there has been insufficient attention paid to capacity building in the industry. The assumption there would be a range of agencies available to offer high-quality, cost-efficient services nationwide has not been borne out in reality. Questions about resourcing and sustainability of agency and workforce development have also been left unanswered. A number of researchers in Australia (Buchanan & Considine 2002; Healy 2004; Meagher & Healy 2003; Owen, Poulos & Eagar 2001; Spall & Zetlin 2004; Wagner & Spence 2003) have identified a range of serious challenges as a result of such changes in community services, including:

- increased workloads and responsibilities (for organisations and their workers) with less funding
- predominantly female employees who are vulnerable to burnout due to low pay
- highly regulated rule-bound jobs leaving little scope for discretion
- workers' talents and achievements not being well rewarded
- limited skills building, with the education and training that workers receive not matching the roles and demands actually encountered on the job.

It is, of course, not a new perception that community services workers are deeply committed to helping clients but that they are under-resourced and under-appreciated. Onyx (1992) reported that the attraction to work in community services was often a values choice (commitment to social values/social change/improving the client group's lot, and/or personal self-development/personal growth).

#### From contracts to partnerships

Further consideration of the social policy literature highlights an ongoing debate about the extent to which a competitive quasi-market model either promotes or constrains good practice. Where funding contracts from government departments to community service agencies have gone to the lowest priced bidder, this has triggered concerns about limited resources for sustainability of the agencies. Initially, some difficulties of managing bids by community service agencies have been addressed by departments providing more information on service delivery expectations, as well as providing assistance to agencies in producing tenders, to enable the agencies to respond adequately to bids. This has not fully resolved concerns of limited capacity building in the community services industry.

Subsequently, there have been moves to use multi-year funding contracts designed to address and stabilise the uncertainties experienced by agencies. In classical contracting, information needs are satisfied by inter-organisational competition: that is, where price is the main distinguishing factor between tenderers the competitive process provides this information. Most service-provision relationships are, however, much better described as relational contracting (Administrative Review Council 1997). Relational contracting (where there is a form of longer term partnership between purchaser and provider) is optimal where the kinds of knowledge possessed by the provider are not widely spread. In such cases, it is more efficient for the purchaser to obtain them through a market-related process rather than to develop and retain them in-house.

Many Australian Government departments have identified a select group of not-for-profit organisations as 'preferred providers', 'core agencies' or 'central' to the mission and operations

of the department. The relationship is more of a 'partnership' than a traditional low-bid contract service negotiation that involves moving from the old 'auditor' model of contract compliance to the new 'partner' model of shared values, expectations, and outcomes. It is, however, a point of debate whether recent developments in partnership relationships between contracting departments and contracted agencies typically strengthens good practice more than previous models of principal-agent contracts and audit models of compliance monitoring.

The change in terminology also does not offer a change to power differentials between the contractor and the provider. There is a huge volume of literature in the UK<sup>3</sup>, and to a lesser extent the US, querying the merits of partnership working, and it is easy to argue that the 'partnerships' currently operationalised in Australia are more rhetoric than reality—or are, as Craig and Taylor (2002) phrased it, 'old rhetoric poured into new bottles'. It has been suggested that the rationale for government versions of partnerships may be primarily based on ideological considerations of reducing welfare dependency and reducing community expectations of the state continuing to take responsibility for ensuring wellbeing (van Gramberg & Bassett 2005). Some agencies insist that this has not resolved difficulties in earlier models of outsourcing, with disadvantages continuing to include:

- costs in adopting new reporting requirements and management information systems
- unrealistic funding to cover service delivery costs including program start-up costs, and reporting demands
- high staff turnover and insecurity for staff who can only be appointed on short-term contracts
- specification of services to be delivered that may not address the agency's mission and core values (Carson, King & Mayer 2007).

Ideally, these developments can increase stability of funding streams for contracted agencies and strengthen community capacity by creating synergies leading to information and skill sharing that enables innovation and streamlining of service delivery (Carson & Kerr 2003). The goal of the relationship then becomes service provision *and* community building. Strengthening the partnership involves the reinforcement of shared values whereby inter-agency networking (building social capital) is complemented by staff learning new ways to collaborate (building human capital). Making these systems work requires considerable investment in staff training, information systems, and contract negotiations, as well as the building of trust among various levels of government and service providers. Successful strategies used by agencies to address these issues include:

- organisational restructuring to reduce overhead and administrative costs
- entrepreneurial capacity building related to fundraising and expanding board competencies

- using other agencies to provide specialist support (financial, human resources, organisational) (Carson, King & Mayer 2007).

This only represents a major advance on early competitive tendering where departments and agencies overcome entrenched practices of maintaining commercial-in-confidence information, and adversarial rather than sharing relationships. Many agencies have limited fundraising options and rely heavily on these service contracts, thus leading them to remain as supplicants or resentful collaborators.

### **Contracting and compliance: implications for evidence-based practice and policy**

Even in situations where the desired outcomes of intervention are broadly agreed—for example that the end product of domestic violence programs is: (1) the increased safety of victims and (2) the cessation of offenders' violent and abusive behaviour—how the interventions are operationalised is shaped and constrained by the funding, service delivery mechanisms and local arrangements. The contracting relationship makes this operationalisation of policy intentions problematic on two levels. It can skew practice and it can skew planning.

It is an empirical matter whether this is to the good or not. First, by funders specifying requirements for service delivery and standardised reporting there is the risk that this can displace emerging practice knowledge. Second, the contractual arrangements can attenuate the links between planning and delivery where the current state of knowledge would benefit from greater integration, particularly in fields like domestic violence where the knowledge about how to achieve desired outcomes is still evolving.

Contracts that stipulate particular service delivery practices or participant/client outcomes usually require a set of data collection and reporting guidelines (Administrative Review Council 1997). These guidelines typically presuppose systematic classification of services and outcomes to enable the production of reports on how much there has been attainment of the required outcomes. Often the data that must be gathered, recorded and reported are those that can be counted and measured in quantitative terms.

Accompanying this requirement is an associated expectation that these data will underpin evaluations that are necessary for evidence-based practice, program and policy even though the descriptive level of individual data items cannot always indicate program success or otherwise. Yet current manifestations of evidence-based practice and policy are said to privilege particular methodological orientations in ways that do not reflect the complexity of epistemological debates about how we identify and understand good practice, and how we can address complex social

problems such as domestic violence. Of course, even in situations where the mechanisms to achieve the desired outcomes involve the application of orthodox knowledge, the mechanisms are not always easy to specify.

It is well documented that service delivery workers (street-level bureaucrats) might modify their practice, interpreting policy and regulations in ways that exert their autonomy or cut corners to minimise stress on the job (Lipsky 1980). In addition, however, there can be at least two other aspects of implementation gaps generated in contracted service delivery arrangements that concern data definition and data analysis.

Evidence-based policy presupposes that there is shared understanding between those collecting data and those utilising the data (Campbell Collaboration 2009). In other words, it is expected that the data being collected are theoretically informed and relevant to understanding the problem at hand, as well as any associated intervention. The specification of data that are to be provided is the province of the funder, but there is potential for gaps at the points of both data generation and data utilisation—whether and how the data are used for developing future policy. It is commonsense that the reporting will presume some consistency in knowledge insofar as the funder and provider have a shared understanding of a theoretical rationale for the data being collected. The process relies on a seamless integration between service delivery and recording ‘outcomes’ but this may not be what happens in practice.

This disjunction between what the organisation delivers and what is recorded and reported may well be at the level of interpersonal and organisational disagreement (politics), but we suggest it may be more fundamental than that. Evidence-based practice and policy presupposes a linear and rational process where there is a shared world view about how to understand the issue, measure it in an appropriate way and then use that knowledge to generate responses to it. This evidence should then be used to inform the contracting of future services. We suggest that while contractualism can use funding agreements to ensure compliance with service delivery protocols and record outcomes in standardised templates, this falls short of ensuring that shared world view.

More than that though, evidence-based practice and policy presumes that data collected and reported will be interpreted in a way that is consistent with that theoretical framework in order to be utilised by agencies to develop a coherent evidence base. On the basis of that coherent evidence base, a theoretically informed analysis by policy decision-makers then constitutes an integral part of the process of developing future policy initiatives. This needs to be a reflexive and recursive process, in that there needs to be observation of practice to assess if these assumptions are valid. At a practical level, this linear approach to policy and practice development

would benefit from stable staffing in policy and practice. While it should not be central to service delivery quality, the rapid turnover of funder and provider staff in the human services also makes the developmental approach to evidence-based policy and practice less likely to occur.

In particular fields of human service delivery, including domestic violence intervention, a search for certainty drives the promotion of evidence-based paradigms and makes assumptions about the appropriate mode of intervention, even though the knowledge base in these areas is still relatively underdeveloped. While practitioners in the field of domestic violence have accumulated relatively comprehensive knowledge of processes at work in the dynamics of this problem, this has not translated easily into forms of intervention that demonstrably reduce the incidence of the problem. Reporting participation in a program (throughput data) is a long way from reporting outcomes, given the state of knowledge about intervention strategies.

### **Variability between agencies: recreating silos**

The dilemmas of coordination within contractual arrangements were highlighted for us recently when we were part of a team that set out to evaluate domestic violence perpetrator programs that were funded by an Australian state government department and contracted out to non-government organisations (NGOs). We investigated service delivery by three community service agencies (one of the agencies operated in multiple metropolitan sites, another agency operated in two sites—one metropolitan and one rural—and the other operated exclusively in a rural site) contracted to deliver men’s programs and associated partner safety contact and support. All agencies were funded via identical contracts, so this gave an opportunity to assess the significance of rural/urban location as well as the importance of organisational culture.

The evaluation design combined three strands: (1) men’s participation and progress through a domestic violence perpetrator program, (2) women partners’ safety and access to support services, and (3) integration of domestic violence responses in the local area. The strands were intended to complement each other notwithstanding the fact that common practice in research on such strands would typically draw on differing conceptual and methodological paradigms.

The literature on contractualism, and the fact that Australian human service delivery has a relatively long history of contractualism, led us to believe that some aspects of the way the programs under investigation were funded and delivered would be of particular relevance to strand 1 of the evaluation. That is, we expected a high degree of standardisation in the agencies’ data collection and program monitoring for the purposes of reporting to the funding department, and expected this to be conducive to the first strand of the evaluation.

At the same time, the literature flags possible disadvantages of contractualism in the human services, namely potential instability in staffing and limitations on agency autonomy and professional practice, including existing inter-agency practices, in ways that might be problematic for strands 2 and 3 of the evaluation.

### **Service delivery variability between agencies**

The agencies were contracted by the department primarily to assess men's suitability for a group program and then run that program. In addition, for women partners of men referred to the programs, the contracts between the department and the agencies had provision for the agencies to contact women partners of the men and offer support. Contracts stipulated input (time allocated to assess men's suitability for a group program; time allocated for victim contact and support; time allocated for running a men's group program of x hours, etc.) rather than output. Reporting to the department consisted essentially of listing membership of groups, and advising where participants did not attend or were disruptive in a group. As a consequence, evaluation of throughput was difficult on the basis of data being routinely provided by the agencies. The information was not sufficient to ensure compliance with the basic requirements of funded service delivery. Thus additional stand-alone data collection has been necessary as a prerequisite for any analysis of output and outcomes.

We observed considerable variation in techniques of assessment of men for perpetrator group programs, and in the conduct of the programs. This amounted to shortcomings in the program logic as well as shortcomings in contract monitoring. The shortcomings were due to the avowed professional model that was purported to inform the practice of two of the agencies (the Duluth model) not being thoroughly operationalised in all agencies. In particular, some sites were not adhering to the logic of the funding associated with outreach to victims and, overall, few women were in contact with the agencies in the course of the men participating in the programs. Thus, aspects of the service delivery did not meet the requirements of the department but, unfortunately, the agencies' data collection and reporting requirements were not sufficiently transparent for the department to be routinely alerted to such shortcomings, and the contracts were not written with sufficiently precise output and outcome indicators for this to be assessed in the course of ongoing quality assurance processes.

Contrary to our initial expectations that a strong standardisation of service delivery would be enforced by virtue of the funding and reporting mechanism, we were surprised to find a much lower level of standardisation in the delivery of the programs than we anticipated in each of the three strands of service delivery, and considerable local autonomy within the existing organisational

culture and procedures of the agencies. This is not necessarily a cause for concern in itself, since variation in practice does not necessarily imply inferior practice, and credibility of a service can be achieved by more than one route. Indeed, where funding provides an opportunity for agencies to tailor variants of a program to be delivered in different sites, this is potentially a positive arrangement where local autonomy and local embeddedness coexist. While standardisation of service delivery and reporting can potentially increase a perception of program quality assurance, and thus credibility, alternatively local embeddedness can also offer a form of local credibility. Differing program interventions is not a problem in itself if they are delivering similar outcomes (increasing victim safety and the cessation of violence and abuse). However, as positive domestic violence outcomes rely, at least in part, on locally integrated responses (Gondolf 2002; Pence & Sheppard 1999), the extent to which the agency is embedded in the local network of domestic violence service responses cannot be as variable as other aspects of the program intervention, as it is critical that they are strongly embedded in the local response.

A defining characteristic of situations we observed was the limited integration and difficulty of coordination between the department and the contracted agencies, as well as between these agencies and other potentially complementary local domestic violence service providers. Despite the fact that criteria for the selection of agencies to be funded included that they had relevant professional experience and embeddedness in the local service networks, this did not ensure that all the agencies were key stakeholders in the local domestic violence community responses. Where we observed limited local service integration, it is important to note that this did not occur across all the agencies. One rural site was clearly offering a service that was well integrated into local support networks for victims, but the established professional networks at that site were more a function of the agency's organisational culture than its rural location, since the same pattern did not exist in the other rural site. Similarly, the differences between the two agencies operating at metropolitan sites were greater than similarities in their integration with their respective metropolitan professional networks.

Where better practice was observed it was a consequence of the organisational culture compensating for and overriding the shortcomings of the contract mechanisms. For example, better practice was identified where workers were motivated by an organisational culture that promoted an existing commitment to ensure victim support in conjunction with men's programs. This was so even where commitment to such standards of practice entailed support from elsewhere in the organisation, and outside the funding for the program. These agencies were able to deliver higher levels of support than those where the driving force

was to deliver only the amount of service that the contract stipulated.

Where there was limited integration it was not, however, just an agency's cynicism in contractual obligations that contributed to that situation. Where limited external integration existed, a key to the different service delivery was whether the funded program had been incorporated into the agency's core business or quarantined and treated as a relatively discrete service delivery arrangement. As a consequence of that basic internal operational decision, where there was a separation of the department's domestic violence funding and service delivery, this limited integration inside the organisation was reproduced and magnified in the external service delivery context.

The consequent short-termism in staffing of this program, with associated high levels of staff turnover, and limited supervision and professional development of group facilitators, exacerbated the internal consequences of program isolation in the external environment. These staffing issues are associated in the literature with increasingly stringent contractual funding arrangements but they need not necessarily exist as a consequence of external funding, as could be seen clearly in one site and to a modest degree in another. If contracts are allocated to agencies with differing approaches, a developmental approach within the agency and between the agency and department can compensate for the variation—but in the absence of careful monitoring of program set-up and delivery then, paradoxically, the circumstances typically associated with over-regulation and monitoring are exacerbated. Thus choices made by agencies that are consistent with their respective organisational cultures need to be assessed more thoroughly before contract letting. And on the basis of that assessment, the funding department needs to treat contractual relationships more as partnerships, in a developmental mode, if there is to be a minimisation of the limited local integration that we observed, and integrated services developed that are consistent with the policy intentions and declared program logic.

### Conclusions

This article has outlined some of the key issues and points of contention in the evaluation of contracted human services. With our work-in-progress evaluation of contracted domestic violence perpetrator programs, we have highlighted some of the implementation gaps in operationalising a dynamic evidence-based approach to improve policy and practice. In our evaluation experience in this instance, it could not be presumed that a standardised contract for services would result in standardised services across a number of sites. In fact, what we found is that it primarily led only to standardised data collection of a fairly basic kind. While the promotion of the standardisation of services has both positives and negatives, in the case of these domestic violence programs a major

concern is that victim safety is not being followed up to the same extent across agencies where this is expected. This raises important issues about contract monitoring by the funder and the funder's recourse to respond to variations in outcomes. This will be influenced by a number of factors such as: the specification of the contract, the extent and type of data available to identify such occurrences, the utilisation of the data by funders, the dialogue with the providers, and the supply/availability of other potential providers in the location should this be a course of action. Since it is a relatively newly developed field, the generation of evidence-based policy and practice is particularly important for domestic violence program development, and thus a commitment to quality service, and operationalisation of that commitment, needs to be better realised in programs and service delivery arrangements such as those outlined here.

### Notes

- 1 This project was funded through the Australian Research Council, Grant LP 0669479.
- 2 A useful caution to this general set of propositions is to remember that unlike governments in other countries, at no time did the Australian Government provide much in the way of community services. Instead, from colonial times it relied on non-government agencies, usually charities and associated not-for-profit NGOs to provide welfare services so that, with the exception of employment services, the growth in this area has not been caused as profoundly by privatisation as it has elsewhere in the world.
- 3 Balloch & Taylor 2001; Glendinning, Powell & Rummery 2002; McDonald 2005; OECD 2001; Osborne & McLaughlin 2004; Powell & Dowling 2006; Stewart 2002; Sullivan & Skelcher 2002.

### References

- Access Economics 2004, *The cost of domestic violence to the Australian community*, Partnerships Against Domestic Violence Program, Office of the Status of Women, Department of the Prime Minister and Cabinet, Canberra.
- Administrative Review Council 1997, *The contracting out of government services: report to the Attorney-General*, Australian Government Publishing Service, Canberra.
- Australian Institute of Health and Welfare (AIHW) 2008, *Linking SAAP, child protection and juvenile justice data collections: a feasibility study*, AIHW, Canberra.
- Balloch, S & Taylor, M 2001, *Partnership working: policy and practice*, The Polity Press, Bristol, England.
- Blay, B & Wright, W 2008, 'No to violence: issues facing men's behavioural change programs', *Australian Domestic and Family Violence Clearinghouse Newsletter*, no. 33, Winter, pp. 18–20.
- Bowen, N, Bowen, G & Ware, W 2002, 'Neighborhood social disorganisation, families, and the educational behavior of adolescents', *Journal of Adolescent Research*, vol. 17, no. 5, pp. 468–490.
- Buchanan, J & Considine, G 2002, *Stop telling us to cope! NSW nurses explain why they are leaving the profession*, Australian Centre for Industrial Relations Research and Training (ACIRRT), University of Sydney.
- Campbell Collaboration 2009, *The Campbell Collaboration: what helps? What harms? Based on*

- what evidence?, viewed 24 June 2009, <[http://www.campbellcollaboration.org/about\\_us/index.php](http://www.campbellcollaboration.org/about_us/index.php)>.
- Carson, E & Kerr, L 2003, 'Theorising state/third sector partnerships: toward a framework of analysis', *Third Sector Review*, vol. 9, no. 1, April, pp. 83–97.
- Carson, E & Wadham, B 2001, 'Regionalism and contractualism: When principles collide', *Just Policy*, no. 24, December, pp. 3–11.
- Carson, E, King, P & Maher, C 2007, *Careers at the coal-face? The community services workforce in South Australia*, Department of Health and Department for Families and Children, Adelaide.
- Craig, G & Taylor, M 2002, 'Dangerous liaisons: local government and the voluntary and community sectors', in C Glendinning, M Powell & K Rummery (eds), *Partnerships, New Labour and the governance of welfare*, Policy Press, Bristol, England.
- Cummings, R 2006, "'What if?": The counterfactual in program evaluation', *Evaluation Journal of Australasia*, vol. 6, no. 2, pp. 6–15.
- Egan, R & Hoatson, L 1999, 'Desperate to connect', *Australian Feminist Studies*, no. 30, pp. 405–414.
- Glendinning, C, Powell, M & Rummery, K (eds) 2002, *Partnerships, new labour and the governance of welfare*, The Policy Press, Bristol, England.
- Gondolf, E 2002, *Batterer intervention systems: issues, outcomes and recommendations*, Sage, Thousand Oaks, California.
- Healy, K 2004, 'Social workers in the new human services marketplace: trends, challenges and responses', *Australian Social Work*, vol. 57, no. 2, pp. 103–114.
- Hendricks, B, Werner, T, Shipway, L & Turinetti, G 2006, 'Recidivism among spousal abusers: predictions and program evaluation', *Journal of Interpersonal Violence*, vol. 21, no. 6, pp. 703–716.
- Klevens, J, Baker, C, Shelley G & Ingram, E 2008, 'Exploring the links between components of coordinated community responses and their impact on contact with intimate partner violence services', *Violence Against Women*, vol. 14, no. 3, pp. 346–358.
- Lipsky, M 1980, *Street level bureaucrats*, Russell Sage Foundation, New York.
- McDonald, I 2005, 'Theorising partnerships: governance, communicative action and sport policy', *Journal of Social Policy*, vol. 34, no. 4, pp. 579–600.
- Meagher, G 2002, *The politics of knowledge in social service evaluation*, UnitingCare Burnside in association with the University of Sydney.
- Meagher, G & Healy, K 2003, 'Caring, controlling, contracting and counting: governments and non-profits in community services', *Australian Journal of Public Administration*, vol. 62, no. 3, pp. 40–51.
- O'Leary, P, Chung, D & Zannettino, L 2004, *A comparative assessment of good practice in programs for men who use violence against female partners*, Office of the Status of Women, Department of the Prime Minister and Cabinet, Canberra.
- Onyx, J 1992, 'Career paths in the community sector', *Nonprofit Management and Leadership* (1996), vol. 6, no. 4, pp. 331–345.
- Organisation for Economic Cooperation and Development (OECD) 2001, *Local partnerships for better governance*, OECD Publications, Paris.
- Osborne, S & McLaughlin, K 2004, 'Modelling government–voluntary sector relations', *European Business Organisation Law Review*, vol. 4, no. 3, pp. 383–401.
- Owen, A, Poulos, R & Eagar, K 2001, *Using the evidence to develop best practice models for identifying initial primary and community care needs: report for the Centre for Health Services Department*, University of Wollongong, New South Wales.
- Pawson, R 2006, *Evidence-based policy: a realist perspective*, Sage, London.
- Pawson, R & Tilley, N 1997, *Realistic evaluation*, Sage, London.
- Pence, E & Shepard, M 1999, 'Developing a coordinated community response: an introduction', in M Shepard & E Pence (eds), *Coordinating community responses to domestic violence: lessons from the Duluth Model*, Sage, Thousand Oaks, California.
- Powell, M & Downing, B 2006, 'New Labour's partnerships: comparing conceptual models with existing forms', *Social Policy and Society*, vol. 5, no. 2, pp. 305–314.
- Shadish, WR, Cook, TD & Campbell, DT 2002, *Experimental and quasi-experimental designs for generalized causal inference*, Houghton Mifflin Company, New York.
- Sherman, LW 1992, 'The influence of criminology on criminal law: evaluating arrests for misdemeanor domestic violence', *Journal of Criminal Law and Criminology*, vol. 83, no. 1, pp. 1–45.
- Sherman, LW, Schmidt, JD & Rogan, DP 1992, *Policing domestic violence: experiments and dilemmas*, Free Press, New York.
- Sherman, LW, Schmidt, JD, Rogan, DP, Smith, DA, Gartin, PR, Cohn, EG, Collins, DJ & Bacich, AR 1992, 'The variable effects of arrest on criminal careers: the Milwaukee Domestic Violence Experiment', *Journal of Criminal Law and Criminology*, vol. 83, no. 1, pp. 137–169.
- Spall, P & Zetlin, D 2004, 'Third sector in transition: a question of sustainability for community service organisations and the sector?', *Australian Journal of Social Issues*, vol. 39, no. 3, pp. 283–299.
- Stewart, M 2002, 'Systems governance: towards effective partnership working', paper presented to the Health Development Agency Seminar Series on Tackling Health Inequalities, London.
- Sullivan, H & Skelcher, C 2002, *Working across boundaries: collaboration in public services*, Palgrave Publications, Hampshire, England.
- Szirom T, Chung, D & Jaffe, R 2003, *National meta evaluation of Partnerships Against Domestic Violence Office for Status of Women*, Department of the Prime Minister and Cabinet, Canberra.
- Trochim, W 2000, *The research methods knowledge base*, 2nd edn, Atomic Dog Publishing Cincinnati, Ohio, viewed 16 January 2009, <<http://www.socialresearchmethods.net/kb/quasiexp.php>>.
- van Gramberg, G & Bassett, P 2005, *Neoliberalism and the Third Sector in Australia*, Victoria University Working Paper, Melbourne.
- Wagner, R. & Spence, N 2003, 'Paucity management practices in Australian non-profit human services organisations', *Third Sector Review*, vol. 9, no. 1, pp. 119–135.