



ISSUES IN

Child Abuse PREVENTION

NATIONAL CHILD PROTECTION CLEARINGHOUSE ISSUES PAPER PUBLISHED BY THE AUSTRALIAN INSTITUTE OF FAMILY STUDIES

NO. 12 AUTUMN 2000

Evaluating child abuse prevention programs

ADAM M. TOMISON

This paper provides an overview of the current state of program evaluation as it is applied to the field of child abuse prevention, the extent to which empirical evaluation has been used, and the degree to which programs have been shown to be effective. Some of the problems that hamper the rigorous evaluation of real world (in situ) applied social programs are discussed. A 'developmental' framework for the comprehensive evaluation of prevention programs is described, as are some alternative approaches that are capable of providing a more flexible response to the demands of evaluation.

In the first half of the twentieth century the patronage of eminent practitioners and academics was commonly deemed sufficient to ensure the development and funding of interventions to improve health and wellbeing and/or to reduce social ills (Crouch 1998). The 1960s heralded not only the modern 're-discovery' of child abuse via Kempe and colleagues' work on the 'battered child' syndrome (Kempe et al. 1962), but also the first empirical tests of the effectiveness of health and welfare programs. These were applied initially to the assessment of generic early intervention programs, such as the Perry Pre-School and Head Start programs in the United States, which were designed to eliminate social and economic class differences by improving the cognitive and social competence of disadvantaged young children (Jacobs 1988; Zigler and Styfco 1996; Ochiltree 1999; Tomison and Wise 1999).

These studies heralded the dawn of the program evaluation era, and with it,

the expectation that public sector programs should be able to objectively and scientifically demonstrate program success and client satisfaction (Rist 1997). This shift has

eventuated, partly, as a consequence of the predominance of economic rationalism and the associated need to demonstrate cost-effectiveness and cost efficiency (Rees 1994; Cooper 1997; McGurk 1997), and ongoing concerns regarding the social cost of poorly performing programs (Weiss 1988; Stevens 1999).

The validity of the latter has been demonstrated with respect to child abuse and neglect, by the continuing toll of child maltreatment deaths (Stevens 1999), high rates of re-abuse and repeated child maltreatment reports to statutory child protection services, and the continuing high incidence of child maltreatment in the community. Curtis (1997) argues that it is 'the seductive appeal of absolute certainty' thought to result from the use of quantitative, economically-focused performance criteria that has led to the

The National Child Protection Clearinghouse serves as an interchange point for information, research and initiatives supporting work in the field of child abuse/neglect prevention.

Coordinator: Judy Adams

© Australian Institute of Family Studies –
Commonwealth of Australia 2000

ISBN 0 642 39473 3
ISSN 1326-8805

Design by Double Jay Graphic Design
Printed by Impact Printing



**Australian Institute
of Family Studies**

300 Queen Street
Melbourne 3000 Australia
Phone (03) 9214 7888 Fax (03) 9124 7839
Internet www.aifs.org.au/
Print Post Approved PP349181/00604

domination of empirical evaluation¹ methods, an appeal that has been strengthened by the absence of other viable alternatives.

The increasingly visible role of empirical program evaluation in public sector management, and the rigorous evaluation of program outcomes in particular, has been described as a relatively recent triumph of 'empiricism over contemplation' (Shils 1980). The incorporation of an adequate program evaluation component is currently a key selection criterion for most funding bodies. Program evaluation and the concomitant shift to greater fiscal accountability (Cooper 1997), has underpinned a general expectation that governments will be able to deliver cost effective, innovative, high performance programs. For them, the pressure is now to work smarter, not just harder' (Cooper 1997:27).

This paper is designed to provide an overview of the current state of program evaluation as it is applied to the field of child abuse prevention, (particularly primary and secondary prevention)², the extent to which empirical evaluation has been used and the degree to which programs have been shown to be effective. Some of the problems that hamper the rigorous evaluation of real world (in situ), applied social programs are discussed. Finally, a 'developmental' framework for the comprehensive evaluation of prevention programs is described, as are some alternative methodological approaches and techniques that are capable of providing a flexible response to the demands of a variety of evaluation models.

DEFINING PROGRAM EVALUATION

Program evaluations are perceived as providing a relatively objective vehicle for quality assurance and a systematic method of data collection and analysis. They may enable an analysis of service utilisation and the profiling of service users; inform ongoing improvement and refinement of program content, provide a measure of overall program success for funding bodies and stakeholders and thus, can inform public policy decision making (Kaufman and Zigler 1992; Willis, Holden and Rosenberg 1992; US Department of Health and Human Services 1995; Nixon 1997). 'A carefully implemented evaluation can move an argument from a discussion of opinions to a review of the evidence' (Pietrzak et al. 1990:10).

There are a multitude of program evaluation types, methods and a variety of terms are used to describe them. Whether the term used is: 'process evaluation', 'formative evaluation', 'program monitoring', 'performance audits' or 'economy and efficiency audits', the intent is the same: Find out what is going on so that responsible decision makers can make the necessary corrections and modifications to keep the program or policy on track' (Rist 1997:35). For the purposes of discussion the 'systems evaluation' model, which is widely used in the program evaluation literature (Pietrzak et al. 1990), will be used to broadly categorise evaluations within this paper into three forms: *input evaluation*; *process or implementation evaluation* and *outcome evaluation*. Although they can be used independently, used together, the three types can provide a comprehensive assessment of program performance.

Input evaluation

Input evaluations, in conjunction with process and outcome evaluations, are effectively an attempt to document precisely the elements of a successful program, or to identify possible flaws or omissions in an unsuccessful program (Pietrzak et al. 1990). Program inputs, such as staffing, client characteristics, resource availability and the role of any ancillary or support services are identified and assessed against the criteria of the community's service needs, service provider goals, best practice standards and cost effectiveness (Pietrzak et al. 1990).

Process evaluation

Also known as *implementation* or *formative evaluation*, process evaluations are designed to investigate program integrity (Nixon 1997) by determining the extent to which a program is operating as intended via the assessment of ongoing program elements and the extent to which the target population is being served (US Department of Health and Human Services 1995).

The aim is to assist service providers to identify areas for change that can enhance service delivery. Usually such evaluations involve the collection of a detailed description of a program's operation and the general environment in which it operates, including the persons served, the services provided and the costs involved (Schalock and Thornton 1988). Key questions are: Has the program or training been implemented as planned?

Has the target population been accessed effectively? Have collaborative links with other programs or service providers been successfully established?

Outcome evaluation

Outcome evaluations, alternatively known as *impact* or *summative* evaluations, are designed to assess the extent to which a program or intervention affects participants on a set of specified outcomes, variables or elements. That is, how has participation in the program affected participants' lives? Any change is assumed to result from participation in the program; the validity of this assumption is tested via comparison between samples in the target population (that is, via the comparison of one or more 'treatment' groups with a 'no treatment' comparison) (US Department of Health and Human Services 1995). Such 'comparison-group research designs are essential in outcome evaluations' (Reynolds 1998:512).

To be successful, outcome evaluations require that a program is well-established and stable: 'program development and outcome evaluation do not mix. If possible, programs should wait until their goals and methods are well-defined and established before initiating an outcome evaluation' (Ellwood 1988:313). The key question is: Are participants exhibiting the expected changes in knowledge, attitudes and behaviour?

Comprehensive program evaluation requires the completion of input, process and outcome evaluations – with input and process evaluations informing the latter (Schalock and Thornton 1988; Pietrzak et al. 1990). Although there is some

acknowledgment of the importance of input and process evaluations for refining service delivery to meet the needs of the identified participant group (and to make processes transparent enough to enable replication), the predominant interest in program evaluations is centred on outcome evaluations and the demonstration of causal relationships between participation in a program and a reduction in social ills or the enhancement of health and wellbeing.

OUTCOME EVALUATIONS

Outcome evaluation research can be classified according to the research design and methodology employed, and by the setting in which the research is conducted. For example, the evaluation may be categorised as either *efficacy research* – where interventions are assessed under highly controlled, laboratory-style, environments involving the provision of intensive professional support to participants. This represents the best case scenario where both service provision and evaluation are conducted under ideal conditions.

While such research may represent the best chance of identifying positive change due to a program of activity, this level of professional support for participants and the level of experimental control available rarely exist in situ (Bickman 1999). In contrast, *effectiveness research* involves conducting evaluations in situ, under actual program conditions and with typical clinicians and ordinary clients (Hoagwood, Hibbs, Brent and Jensen 1995). Any loss in experimental control is balanced by the increase in ecological validity of demonstrating the effectiveness of actual programs in operation.

The US National Committee on the Assessment of Family Violence Interventions (Chalk and King 1998) used the degree of experimental rigour or strength of evidence present in an evaluation to produce a hierarchy of outcome evaluation designs for which there is general consensus. That is, the extent to which specific evaluation designs can determine the unique effects or impact of a prevention program or other intervention, beyond any change that may have occurred because of any other factors.

Non-experimental designs

At the lowest level of evidence in the outcome evaluation hierarchy are non-experimental designs, simple pre- and post-test analyses, case studies, anecdotal reports, or client feedback and satisfaction measures. Such studies may produce useful information for service providers (or researchers developing research), such as a profile of the characteristics, experiences and presenting problems of participants, or enable the identification of program implementation or process issues. In spite of their utility as 'important building blocks' (Chalk and King 1998), little control is able to be exerted over the program environment and the resultant data is insufficient to accurately determine causation or program impact.

Quasi-experimental research

Depending on the criteria employed when assessing program evaluations, quasi-experimental research can be

considered as the lowest acceptable form of evidence (such research may be deemed as unacceptable by the methodological purist) when assessing program impact. Such research typically involves at least two groups of participants. First, the *treatment* or *experimental* group – individuals who participated in the evaluation; and second, some form of *comparison* group, often taken from program waiting lists – involving individuals comparable in character and experience who have not received the intervention, or who have received a different intervention (Fink and McCloskey 1990; Chalk and King 1998).

Alternatively, the treatment group may be assessed a number of times – prior to the intervention, during intervention and post-intervention, in order to assess if participation in the program significantly changed some aspect of the participants that was maintained over time; this is known as a *time series* design (Campbell and Stanley 1966). Such designs provide evidence of the existence of a relationship between program participation and particular outcomes. Studies with a comparison group, or that involve a time series methodology are generally perceived as more reliable than a simple pre/post-test design. However, because the comparison group is not assigned randomly, it is not possible to derive unbiased estimates of the magnitude of the effects produced specifically by a program.

Experimental designs

The highest level of evidence is derived from experimental designs, with randomised control trials (RCTs) considered as the 'gold standard' or best practice evaluation design (Campbell and Stanley 1966; Fink and McCloskey 1990; Nixon 1997; Chalk and King 1998; Smith 1999). Such designs incorporate the strongest degree of experimental control over the program environment, restricting a number of threats to internal validity and dealing with extraneous variables. The major difference between quasi-experimental and experimental studies is that the latter involve the random allocation of participants to either an experimental group or a control group, enabling the researcher to overcome a variety of potential sources of bias and provide the best chance of determining an unbiased estimate of the effect of participation in a program (Chalk and King 1998). Such designs are often referred to as scientifically or methodologically rigorous, or as the 'true experiment' (Fink and McCloskey 1990; Chalk and King 1998).

CURRENT KNOWLEDGE

In spite of the vast number of program evaluations that have been performed on a variety of child abuse prevention programs, and the general acceptance that scientific evaluation should be an essential part of all prevention programs, very few rigorous evaluations have been done in Australia or internationally (Fink and McCloskey 1990; Harrington and Dubowitz 1993; James 1994; Melton and Flood 1994; Tomison 1995; 1997a; 1997b; 1998; Chalk and King 1998). The majority of current evaluations are non-experimental and thus provide no firm basis for determining the relative impact of a program or intervention, or of examining the impact of a particular program or activity on specific

populations (Chalk and King 1998). Similarly, little is known about which interventions are most effective with each sub-population of child abuse and neglect because most studies do not differentiate between types of maltreatment (Whipple and Wilson 1996).

In describing primary and secondary family support and education programs, Weiss described the field as 'a set of two dozen or so flagship research and demonstration programs' [key large-scale, longitudinal RCT studies] and a 'larger grass roots fleet of small and fledgling community-based programs. Both the fleet and flagship programs have had uncertain credibility and funding and little visibility, particularly among policymakers and human service professionals' (1988:3). Subsequent meta-analyses of evaluation research have reached similar conclusions (Hanson 1997).

International studies

MacDonald and Roberts (1995), in their review of child abuse prevention programs in the United Kingdom, commented that the vast majority of interventions had not been evaluated prior to introduction, and to all intents and purposes had the status of uncontrolled experiments. Fink and McCloskey (1990) reviewed 13 United States program evaluations recommended by experts and published from 1978 to 1988. Using the 'true experiment' as their criterion, they reported that most of the studies were methodologically sound. That is, the studies had control groups or involved longitudinal research enabling an assessment of program effects on families over time.

However, Fink and McCloskey concluded that the evaluation studies were hampered by a lack of uniform definitions of child maltreatment and of what constituted an 'at risk' child or family, that the studies had not fully measured the impact of programs on the incidence of child abuse and neglect and had failed to collect data on some of the indicators that were targeted for special attention in the prevention programs themselves. They concluded that it was not possible to determine whether specific aspects of family functioning had improved as a result of participation in the project.

US national assessment

Currently one of the best resources available for assessing the effectiveness of family violence prevention and treatment programs is the US National Committee on the Assessment of Family Violence Interventions report developed for the US National Research Council (Chalk and King 1998). Overall, the committee concluded that evaluation studies 'are usually small in scale, likely to be underpowered, and subject to a long list of rival interpretations' because of flawed or limited study designs and methods of assessment (Chalk and King 1998:91).

The committee only identified 114 evaluation studies conducted in the period 1980–96 which were of sufficient methodological rigour (that is, an experimental or quasi-experimental investigation of program effectiveness that incorporated a control or comparison group) to enable inferences to be drawn about the effectiveness of specific interventions in the area of child maltreatment and other

family violence (Chalk and King 1998). Overall, 78 of the 114 evaluations focused on child maltreatment interventions. They noted that it was unclear whether the lack of rigorous evaluations was due to a lack of resources, the limited duration of studies which prevented the collection of a sufficiently large sample for analysis, or the lack of pre-evaluation research that described service operation and the nature of the preventative intervention, thus providing a foundation for detailed evaluation. As a consequence however, there is currently a dearth of evidence on 'what works, for whom, and under what conditions' (Chalk and King 1998:91).

Home-visiting programs

An exception to the rule in terms of child abuse prevention programs, was home-visiting services, which were found to be very effective in detecting and identifying maltreating families and/or alleviating concerns once the cases are 'known' (Olds et al. 1986a; Olds et al. 1986b; Nelson, Saunders and Landsman 1993; Olds et al. 1997; Chalk and King 1998). The most scientifically rigorous program evaluation of a comprehensive home-visiting program, and arguably one of the most rigorous evaluations of a child abuse prevention program, is the Prenatal/Early Infancy Project developed by David Olds and colleagues (Olds et al. 1986a; Olds et al. 1986b; Olds et al. 1997). This frequently cited project has been described in detail in a number of previous Clearinghouse publications (for example, Tomison 1998, Tomison and Wise 1999).

Australian research

With a few exceptions, no systematic research has preceded the implementation of primary and secondary prevention programs in Australia. In many cases, overseas programs have been adapted for use without any investigation of the needs of the community for which the program was intended (James 1994; Tomison 1995; Vimpani, Frederico, Barclay and Davis 1996) – although this appears to be changing (Tomison 1997b). Given the identified limitations of international program evaluations and knowledge of program effectiveness, the Australian reliance on international evaluation studies and the efficacy of international programs may be misplaced.

James (1994) in her audit of the National Child Protection Clearinghouse *Prevention Programs* and *Research* databases, noted the distinct lack of attention paid to the evaluation of family support/parenting programs in Australia, stating that 'none have been effectively quantified in terms of actually measuring reduction in the incidence of child abuse and neglect' (1994:3), despite the quite extensive use of such programs across the nation. Much like Fink and McCloskey (1990) and MacDonald and Roberts (1995), James (1994) concluded that despite some promising evaluation research, there was a need for better quality evaluations that incorporated more methodologically rigorous designs.

In a subsequent audit of the National Clearinghouse *Prevention Programs* database however (Tomison 1995), it was apparent that service providers had become more aware of the need to implement program evaluations with some

degree of methodological rigour. Although the majority were still non-experimental designs, a sizeable proportion of programs were reported to incorporate a pre-test/post-test evaluation design, which was seen as a positive development.

New South Wales audit

In 1997 the National Clearinghouse undertook an in-depth assessment of 453 child abuse prevention programs operating in the State of New South Wales for the NSW Child Protection Council (Tomison 1997b). As part of this state audit, an assessment was made as to the nature and extent of program evaluation adopting the criteria used in a Canadian audit of 'social prevention' programs (Dallaire, Chamberland, Cameron and Hébert 1995). In this project 27 different program effectiveness criteria were synthesised down to five global factors, according to the form of assessment that was being undertaken and the extent (and type) of participant feedback:

- *Participation and satisfaction* – typically quite simple evaluations based on an assessment of client feedback (user satisfaction) and attendance at sessions. [It should be noted that often prevention programs are parent rather than child-focused; thus, measures of client satisfaction and attendance are often parent-focused. It is important to acknowledge that parental perceptions of program 'success' or positive change may not necessarily be experienced as such by the child (Butler and Williamson 1994)];
- *Effects on users (Outcome evaluation)* – the extent to which there are measurable changes in clients as a result of participation in a program. Evaluations are generally of a pre-/post-test design, where the measures employed range from simple knowledge questionnaires to psychometric tests and detailed behavioural observation;
- *Characteristics of the agency (Process evaluation)* – an assessment of observable changes in the structure and organisation of the host agency which reflect the viability of the program (for example, budgetary increases, expansion of the program);
- *Characteristics of the intervention (Process evaluation)* – a quality assessment of the program, which usually includes a determination of the generalisability of the program (for instance, its adoption by other agencies) and a determination of whether the program is adequately accessing the target population;
- *Neighbourhood changes* – an evaluation of the extent to which the program results in identifiable changes to the local community, such as the increase or decrease of a social problem (identified via an assessment of health and well-being indicators)³, or the development of new community groups or structures as a result of a program (such as the creation of self-help groups).

Overall, approximately one-third of programs included in the Audit did not conduct any form of evaluation. A further 10 per cent failed to provide sufficient information to enable an assessment of their evaluation (for example, noting only that 'the program was a success', or 'an evaluation is planned for later in the year'). Of the remaining programs (categories not mutually exclusive), approximately 80 per cent (199 of

the 248 programs that supplied details of an evaluation) based their evaluation on simple client (mainly parent) satisfaction measures and levels of attendance (Participation and satisfaction). In a discussion of the evaluation of parent education programs, Hobbs et al. concluded 'evidence of accessibility, use, and [participants'] satisfaction is perhaps all that is possible and all that should be required. That evidence is available, and it is uniformly positive' (1984:267).

However, 34 per cent of the evaluations did involve a pre-test/post-test design used to assess a program's effect on clients – the vast majority without a comparison or control group (Effects on users – outcome evaluation); 40 per cent focused their assessment on the effectiveness of the program at attracting participation from targeted populations and/or the generalisability of the program (Characteristics of the intervention); and 19 per cent of the evaluations attempted to document changes to the host agency that resulted from running the prevention program (Characteristics of the agency). While *characteristics of the intervention or agency* may be important facets of a comprehensive systems evaluation (Pietrzak et al. 1990), in isolation they do not constitute evaluations of the extent to which the program is able to reduce the likelihood of child maltreatment and/or promote positive child and family relationships.

Implications

It was apparent from the Audit that there had been an increasing acknowledgment by service providers of the importance of assessing program performance. From an empirical perspective it was apparent that a number of agencies had attempted to incorporate a degree of methodological rigour in the evaluation performed, although none could be classified as 'rigorous'. Given the previously mentioned limitations of applying empirical designs in situ (see the next section for further discussion), and the degree of expertise required to conduct such outcome evaluations, this is not a surprising result. Particularly when the majority of the programs were being run by grass roots agencies or groups (the so-called 'fleet' programs, Weiss 1988), who often had neither the financial nor personnel resources that would enable such an approach. Thus, many of the evaluations appeared to have been conducted more as a means of program refinement, to fine tune service provision (process evaluation), than as an attempt to demonstrate a causal relationship between program participation and long-term outcomes (that is, reductions in child maltreatment and other social ills).

Not surprisingly, it appeared that it was the larger non-government service provider agencies, government services (such as hospitals, regional health services or the state education department), and university-supported projects, that had more success at developing outcome evaluation designs that approached methodological rigour. These agencies either had the resources and staff expertise to undertake such an evaluation, or were able to fund an independent evaluation by external consultants.

Some of the larger agencies also had an additional option available when considering the implementation of a program evaluation. When a specific prevention program was operating across a number of individual agency centres, (for

example, the Family Work Program run by various Burnside centres), the data could be aggregated to provide a greater quantity of information for the evaluation than could be produced at one site, thus strengthening the power of the evaluation to provide useful results (Tomison 1997b). To ensure such a multi-site evaluation was valid, required a standardised program implementation process, the adoption of uniform data collection methods across the sites and the adequate experimental control of any inter-site differences that might affect service delivery (for example, the client population, staff training and experience).

In summary, the findings indicated increased acknowledgment of the need to conduct evaluations and clear attempts to apply some degree of rigour to the assessment of program 'success'. The Clearinghouse is currently analysing data from a recent Australian National Audit of child abuse prevention activity which will enable a firmer picture to be developed of the nature and extent program of program evaluations that are currently being conducted by service providers across the country.

PROBLEMS WITH EMPIRICISM – APPLYING SCIENCE IN SITU

As is demonstrated by the evidence summarised above, although rigorous experimental designs may be best suited to ruling out competing explanations for observed effects and to indicate causal relationships between programs and subsequent outcomes (Devine, Wright and Joyner 1994), there has been a general lack of quality empirical evaluation studies in the child abuse prevention field (Bickman 1992; Hoagwood 1994; Chalk and King 1998). What factors have hampered the conduct of rigorous evaluations?

First, some agencies fear that an evaluation, particularly an external evaluation, may produce negative results or indicate that their program is not successful. Although it is generally the case that any negative findings are balanced by positive effects, it is equally important to discover if a program or some of its elements are not working in order to refine the program and increase the potential for an effective intervention (US Department of Health and Human Services 1995). A failure to evaluate means that a program is operating without clear evidence that it is effective.

Second, it is apparent that although agencies and groups' attitudes to evaluation have changed somewhat, some service providers perceive program evaluation as diverting badly needed resources away from service provision, thus potentially harming participants and placing an additional burden of work on frequently overloaded staff (US Department of Health and Human Services 1995). Yet determining the effectiveness of a program and enabling the refinement or modification of service delivery will benefit participants who may not be receiving the services they need.

Third, the potential complexity and technical nature of evaluations, combined with a lack of in-house expertise and knowledge as to who can provide assistance (US Department of Health and Human Services 1995; O'Donoghue 1997), may often lead service providers to conduct only a cursory evaluation, or to fail to conduct any evaluation

(US Department of Health and Human Services 1995). Fourth, evaluators have to contend with an inability to control important familial and program-related variables, the lack of a uniform standard for program success (Vimpani et al. 1996; Chalk and King 1998; Hutchinson 1999) and:

'significant methodological logistical problems [including] difficulties in constructing and gaining access to appropriate sample sizes, limited availability of comparison and control groups, weak research measures and survey instruments, short time intervals for follow-up studies, and high attrition rates in both the interventions themselves and the evaluation studies. Some interventions that are in widespread practice have *never* been evaluated . . . whereas extensive attention has focused on a few interventions in specialized settings' (Chalk and King 1998:15).

In addition, the realities of service provision are that:

- programs frequently involve the targeting of complex social issues;
- research is typically affected by a multitude of extraneous factors that are difficult to control; programs are multi-faceted and often change to meet the needs of clients. Variations in the components, duration and intensity of treatment and length of follow-up, that is, a program where implementation is unstable, have confounded efforts to identify promising interventions (Bickman 1992; Farmer et al. 1997; Chalk and King 1998);
- inadequate resources and/or measures affect the ability to implement a rigorous evaluation design. Empirical assessment requires the development of a finite set of outcomes that can be accurately measured, within a frequently short evaluation timeframe (Bickman 1992; Farmer et al. 1997; Chalk and King 1998). However, evaluators may be committed to assessing multiple outcomes tailored to the needs of service users (for example, goal attainment scaling – Pietrzak et al. 1990), involving the development and incorporation of specific, non-standard measures to evaluate success. These may impact negatively on the extent to which a rigorous evaluation can be completed (Jacobs 1988);
- problems in conducting rigorous evaluations have been exacerbated by the relatively low priority (until recently) given to prevention programs by governments and other institutions, and the common tendency of funding only short-term demonstration or pilot projects and program evaluation efforts (Melton and Flood 1994; Tomison 1997b).
- empirical evaluations require a high degree of skills, significant resources and often, a long timeframe, when most child abuse prevention programs do not continue to run past a three year pilot implementation (Nixon 1997); and finally
- service providers and evaluators must walk a fine line between methodological and ethical issues, such as the assignment of participants to a 'no treatment' control

group and the subsequent lack of service provision (Chalk and King 1998).

In the following sections the issues of randomisation and the evaluation of 'whole of community' approaches are used to demonstrate the difficulties of attempting to apply experimental control in applied settings.

Randomisation

Despite the empirical imperative to randomly assign participants to either experimental or control groups, random allocation creates the potential for participants who may be in serious need of intervention to be randomly allocated to the 'no treatment' control group. Many service providers will not agree to methodological considerations taking precedence over clinical and ethical considerations in service provision such that families requiring significant service provision are assigned to a 'no treatment' control group (Jacobs 1988; Bickman et al. 1997; Farmer et al. 1997).

Thus, the use of waiting list participants as a matched comparison group in a quasi-experimental study (that is, an accidental sample where participants are assigned to 'treatment' or 'no treatment' conditions as a function program vacancies) is generally seen as preferable by service providers (Bickman et al. 1997; Farmer et al. 1997). In addition, with the shift away from the traditional view of children and families as 'objects' of study, randomisation has also been seen as removing or reducing participants' freedom of choice to make an informed decision to participate (Heflinger 1987).

'Whole of Community' programs

Community-level interventions, such as 'whole of community' programs, the most recent generation of a long line of complex, ecologically-based, community-level interventions (Kahn and Kamerman 1996; Pawson and Tilley 1998), represent the greatest challenge to attempting to conduct empirical evaluations. They provide a clear example of program development outstripping evaluation theory and practice (Swerissen 1999), and present an opportunity to highlight the current role of experimental approaches in evaluating complex in situ programs.

'Whole of community' approaches, better known in the United States as 'comprehensive community initiatives' (CCIs) (Kahn and Kamerman 1996), are comprehensive, multi-level prevention efforts that typically maintain an individual or family-level component, but which also address the socio-cultural context within which children and families live and attend to the various factors underlying social problems like child maltreatment (Hayes and Bowes 1999; Reppucci et al. 1999; Tomison and Wise 1999). Although these initiatives may take a variety of structures and forms, they all have the aim of empowering community members to participate in a partnership with government and the professional sector as a means of promoting the development of healthier communities. That is, they promote positive change in disadvantaged neighbourhoods for individuals, families and the community as a whole, by improving physical, social and economic conditions (Kubisch et al. 1998).

The US Committee on the Assessment of Family Violence Intervention noted that 'the emerging emphasis on integrated, multifaceted, community-based approaches to treatment and prevention services, in particular, presents a new dilemma in evaluating family violence interventions; comprehensive interventions are particularly difficult, if not impossible, to implement as well as study using experimental or quasi-experimental designs' (Chalk and King 1998:59), an assertion supported by a number of other researchers (Hollister and Hill 1995; Midford and Boots 1999). The ability to conduct a rigorous evaluation is hampered by various design and constituent elements of community-based approaches (Chalk and King 1998). These include:

- their comprehensive nature makes determining the theoretical underpinnings of the programs, and thus, identifying the program objectives and rationale for operation difficult;
- difficulty in isolating the effects of the new approaches, given that in many communities the programs often complement or extend pre-existing services, rather than replacing them entirely;
- the programs often explicitly take an holistic approach to service provision, resulting in families receiving a multidisciplinary package of supports, making it very difficult to determine which service, if any, contributed to the clients' improved wellbeing;
- if everyone in a community receives some form of service, it becomes difficult to make a comparison with a 'no treatment' control group;
- if the sequence of program implementation is partly determined by clients' needs, it becomes difficult to distinguish between the effects of any selectivity bias and program effects; and
- the variation between individual communities and the diversity of organisational approaches that may be employed to meet the needs of particular communities impedes any analysis of the implementation of the project or its constituent elements (process evaluation).

Implications

Overall, what is apparent is that 'the stark realities of providing services to families conflict with the ideals of experimental realities' (Vimpani et al. 1996:36). The current trend of framing service delivery in terms of cost effectiveness and efficiency (Rees 1994), (and the difficulties associated with addressing the issues described above via traditional rigorous evaluation), has meant that 'not only are programs, services and systems of care under pressure to demonstrate effectiveness, but so too are our evaluation methods and procedures' (Nixon 1997:2-3).

Given the difficulties inherent in conducting methodologically rigorous research, why is empirical evaluation still promoted? Is it realistic to judge evaluation research by the 'gold standard' of scientific rigour?

WHY EMPIRICISM?

In spite of the difficulties inherent in conducting empirical studies in situ, their limited ability to demonstrate causality with what are often complex programs designed to address complex social issues (Pawson and Tilley 1994; Hawkins and Leigh 1997) and the many questions such evaluations may leave unanswered, there are still many committed advocates of the 'gold standard' of empirical evaluation (Halbrook et al. 1997; Hawkins and Leigh 1997; Chalk and King 1998; Hutchinson 1999). Weiss reflects the great hope of empirical evaluation noting that 'if strong program and policy-relevant evaluations can be designed and implemented, family support and education programs may not evolve into just another short-lived and faddish panacea for social ills. Instead, they may be able to serve as central building blocks for a human service system realigned around prevention and the promotion of family health and well-being' (1988:4).

Because empirical research has perhaps the greatest potential of current methods to demonstrate a causal relationship between participation in a program and subsequent positive outcome (Nixon 1997), some researchers maintain the belief that greater efforts need to be expended to ensure such studies are completed, in spite of the various hurdles that make such studies exceedingly difficult to achieve (Halbrook et al. 1997; Hutchinson 1999). Epitomising the empiricist position, Stevens contended that 'using large samples, control groups and looking at strictly defined aspects of experience which can be measured is the only way of providing strong evidence about outcomes or any other aspect of human life. Other methods are subservient' (1999:30).

Thus, despite acknowledging the difficulties inherent in conducting large scale empirical outcome evaluations, even by 'highly skilled researchers from very capable organizations', there remain many advocates who believe that 'problems notwithstanding, more controlled intervention research is clearly needed' (Fraser, Nelson and Rivard 1997:148). It has been posited that the continued emphasis on empiricism common in social science research and medicine is really a mid-20th century phenomenon (Crouch 1998). That it is still supported in the face of its limitations, is thought to emanate, in part, because of the influence of older generations of researchers who perpetuate an empiricist approach through their senior academic and professional roles, as 'gatekeepers' of academic institutions (Mishler 1979; Crouch 1998).

The difficulties of utilising empiricist approaches is also evident in medical research, with 'most research published in medical journals [being] too poorly done or insufficiently relevant to be clinically useful' (Godlee 1998:6). Godlee cites a medical survey (Haynes 1993) which reported that 'more than 95 per cent of articles in medical journals failed to reach minimum standards of quality and clinical relevance' (1998:6).

In spite of such claims, empiricism and the RCT still reigns supreme over medical research where 'any omission of statistical rigour from clinical research is viewed as an act

of heresy, and, like the ideas of the heretic, the work of the insufficiently rigorous is banished from the scientific literature' (Crouch 1998:1101).

But are randomised control trials really an effective tool? Are the additional benefits they may produce worth the difficulties of applying the method to the study of complex, real world programs and social issues?

The benefits of randomised control trials

In 1999 McKee and colleagues made a comparison of 18 studies where an intervention had been evaluated by both RCT and non-RCT methods and concluded that while the magnitude of treatment effects differed depending on the method used, neither method gave consistently greater effects than the other. Thus, differences in effect sizes could not be assumed to be solely due to methodological differences (that is, the presence or absence of randomisation) as characteristics of the intervention, study population and/or participants' preferences could all impact on the results (McKee, Britton, Black, McPherson, Sanderson and Bain 1999).

Similarly, even with careful matching and random assignment, differential rates of participant dropout (that is, withdrawal from the program) in the treatment and control groups, may lead to significant group differences on important variables (Miner 1997). Rather than dismissing quasi-experimental research, McKee and colleagues concluded that: 'RCTs and non-randomised studies can provide complementary evidence – but it is important that clinicians using this evidence are aware of the strengths and weaknesses of each method' (McKee et al. 1999:315).

Some however, have argued that the issue for empiricism is less about the added benefits of randomisation, but more about the value of empiricism as a whole for the assessment of in situ social programs. Pawson and Tilley enquire: 'What is it about *quasi-experimental* research which leads even the very best of it to yield so little' (1994:297).

The limitations of empiricism

First, above and beyond the problems inherent in adopting empirical and quasi-empirical designs in situ, the efficacy of empirically-based outcome studies are further limited by a failure to delineate the programmatic elements that have led to 'success':

'Even if skeptics concede that the . . . treatment was effective, the active ingredients would remain unknown. Was it due to program content? The therapy process? The setting? The skills or personalities of the staff? Or to any number of other potentially relevant factors? Given that most treatment effects are unlikely to be strong, skeptics may simply dismiss any positive . . . findings as an anomaly, in much the same way that skeptics disregard statistically significant results from ESP experiments' (Hanson 1997:133).

Environmental influences

Second, fundamental to empirical research is the assumption that phenomena exist in pure, variable forms,

uncontaminated by environmental effects (Jacobs 1988). This assumption does not hold in situ, where ecological (socio-environmental) influences have been demonstrated from the 1930s onwards (Bronfenbrenner 1979). Thus, conducting empirical research requires an attempt to control extraneous environmental influences, to hold these factors constant during the intentional manipulation of the variable under examination in an attempt to replicate the purity of the laboratory. Such an approach may represent a simplistic transformation of a program into statistics that are insufficient to represent the reality of the program's operation (Guba and Lincoln 1981; Wadsworth 1982; Pawson and Tilley 1994) and result in a cumbersome, narrow evaluation of limited validity and utility (Jacobs 1988). The Westinghouse Learning Corporation evaluation of the US Head Start program in the late 1960s represents an early example of the limitations of empiricism (Jacobs 1988).

Head Start was based on a 'whole of child' philosophy, and embraced a number of goals. Each Head Start preschool centre was designed to improve children's mental and physical health, enhance their cognitive skills, encourage their social and emotional development, self-confidence, healthy relationships and social responsibility, and foster a sense of dignity and self-worth for both the child and family. Overall, the primary aim of the Head Start program was to develop children's overall social competence through the comprehensive provision of services. The most unique aspect of the project was the central role accorded to parents, who were encouraged to be involved in the planning, administration, and daily activities of their local centres (Zigler and Styfco 1996).

Despite the program's aim of improving both social and cognitive competencies, the empirical Westinghouse Learning Corporation evaluation was centred on a few standardised measures of intelligence and achievement. The findings indicated that previously identified increases in the child participants' intelligence scores, which were thought to be potentially stable effects, were actually only short-term benefits which dissipated quite quickly. It has been contended that these findings contributed to first, widespread disillusionment and disappointment among the proponents and supporters of Head Start, in some cases a belief that poverty was an intractable social ill (Jacobs 1988), and third, a subsequent erosion of public support (albeit temporary in nature) for the US early childhood agenda (Weiss 1983). Further careful evaluation, tailored to the Head Start objectives, subsequently identified a variety of other lasting program effects, such as children being less likely to fail a grade or to require special education classes, benefits which lasted until many children had reached the age of 12 years or more (Zigler and Styfco 1996). There was also strong parental approval for the program and the effects it had had on their children (Ochiltree 1999).

The Westinghouse findings have been used to highlight the limitations of applying science to the study of complex social programs, the dangers of reducing complex programs to a few scores, and the need to ensure that outcome measures are able to effectively measure the variety of objectives typically outlined in social programs. The study reinforced the need to determine not only if a program will

'work', (often such programs will lead to change in some children and/or families), but more importantly, why the program works, for whom and in what way (Jacobs 1988). Such 'big failures' (Jacobs 1988) enabled the adoption of more qualitative-descriptive methods (Patton 1980; Firestone and Herriott 1983) and a gradual move away from a primary outcome-oriented function; the 'systems evaluation' model (Pietrzak et al. 1990), was developed as a result of such early rigorous evaluations.

Resource issues

Third, prior to conducting a rigorous examination of effectiveness, prevention programs require adequate, longitudinal funding and sufficient start-up time to address procedural problems and to ensure stable program implementation (Nixon 1997). Given the current trend of funding predominantly short-term pilot or demonstration projects (Tomison 1997b), and a lack of in-house research expertise in many grass-roots service providers, many will not have articulated a set of goals and/or will be too inexperienced or overworked to collect anything above basic service utilisation data (Jacobs 1988).

Therefore, while it may be appropriate to expect that rigorous experimental or quasi-experimental evaluations are carried out by large, complex organisations operating prevention programs, it is unrealistic to expect that the majority of evaluations (the 'fleet' evaluations – Weiss 1988) will be able to develop an outcome study at anywhere near the level of the 'gold standard'. The adoption of a degree of rigour may enhance any evaluation, expectations must however be tempered by taking into account the level of resourcing, knowledge and expertise available to the grass roots agencies or groups that run the majority of prevention programs in the US and Australasia (Weiss and Hite 1986; O'Donoghue 1997). As one service provider noted: 'most of the small efforts made within community agencies are so questionable and easily challenged, however they do provide us with information. We must make do with what we can get rather than PhD research studies' (King 1998:5).

Some researchers (Parker, Ward, Jackson, Aldgate and Wedge 1991; Nocon and Qureshi 1996) have contended that it is perhaps more realistic to look for 'patterns of benefit and loss' when attempting to establish outcomes in both social care and/or child welfare, rather than to attempt to produce a uni-causal model (Frost 1989).

Replication issues

Fourth, even exceptionally well-designed empirical evaluations may have an important failing: such longitudinal, prospective outcome studies are exceedingly difficult to replicate. Hanson (1997) provides a description of the California Department of Mental Health's Sex Offender Treatment and Evaluation Project (SOTEP), (a long-term prospective outcome study of sex offender treatment), a study that was 'exceptionally well-designed, but its methodology has one important failing: it is almost impossible to replicate (Hanson 1997:133).

To conduct rigorous outcome studies require a significant commitment of time, funds, service provider cooperation and a 'rare confluence of research talent and political will

that is unlikely to be repeated in the foreseeable future' (Hanson 1997:133). The opportunities to conduct the research are becoming fewer, despite the continued promotion of rigour. There has been a clear retreat away from large-scale assessments and evaluations of national demonstration projects and independent evaluations to internal evaluations, particularly in Canada and the US where there had previously been financial support for large, long-term, expensive evaluations (Rist 1997). Why has this occurred?

Difficulties in operationalising objectives and measures, the prohibitive cost of large-scale evaluations (smaller is cheaper); and the shift away from attempting to develop societal-level alternatives for social problems (and thus large-scale evaluations of these strategies) have all impacted negatively on funding bodies' willingness to finance the big studies. Rist argues that causation is no longer the driving force in evaluation research. Rather, there is now more of a focus on organisational performance and the potential for evaluations to produce continual improvement (Rist 1997).

In addition, that decades of evaluation research has produced very few scientifically useful conclusions has led to a reluctance on the part of service providers and evaluators to claim that a program or policy can lead to direct and explicit behavioural outcomes (Rist 1997). Evaluators have become much more realistic about what evaluations may accomplish and how they may be used (Rist 1997; Little 1999); program effects are now seen as being more likely to be 'indirect, long-term and cumulative' (Rist 1997).

Yet concomitantly, the expectations of policymakers as to what may be accomplished by evaluation are growing. They have the expectation that evaluations can help address issues of accountability and improved performance while the time available for an evaluator to respond to political needs for information is growing shorter (Rist 1997). The growing gap between expectations and performance may have resulted in a growing dissatisfaction with costly, prospective studies that are, at the end of the day, unable to provide the surety of 'fact' demanded by funding bodies.

Overall then, while there may always be a place for empirical 'flagship' evaluations with a broad outcome focus that may 'prove' a program's effectiveness, there has been some recognition that experimental rigour in isolation, is often an unsuitable means of evaluating social programs. This has become more apparent as service providers have adopted complex, multifaceted, ecological approaches to addressing or preventing what are complex social issues (Hollister and Hill 1995; Chalk and King 1998; Pawson and Tilley 1998; Midford and Boots 1999). As Pawson and Tilley contend:

'it is high time for an end to the domination of the quasi-experimental (or OXO) model of evaluation. Such an approach is a fine strategy for evaluating the relative performances of washing powders or crop fertilizers, but is a lousy means of expressing the nature of causality and change going on within social programmes' (Pawson and Tilley 1994:292).

If it is assumed that empiricism should only play a limited role in the evaluation of social programs, what are the alternatives?

A COMPREHENSIVE EVALUATION FRAMEWORK

The US National Committee on the Assessment of Family Violence Interventions concluded that:

'the field cannot be improved simply by urging researchers and service providers to strengthen the standards of evidence used in evaluation studies. Nor can it be improved simply by urging evaluation studies be introduced in the early stages of planning and design of interventions. Specific attention is needed to the hierarchy of study designs, the developmental stages of evaluation research and interventions, the marginal role of research in service settings, and the difficulties associated with imposing experimental conditions in service settings' (Chalk and King 1998:60-61).

The basis for any alternative to empiricism is the development of a comprehensive evaluation framework that can enable service providers to make the most of their resources and exploit any evaluation opportunities. A comprehensive evaluation requires the development of an understanding of how a program's structure influences the process of service delivery and service delivery influences outcome (Donabedian 1978, as cited in Pietrzak et al. 1990). Information is provided on not only a program's level of effectiveness, but also the *reasons* for its effectiveness (Pietrzak et al. 1990).

Despite a general focus by many researchers, governments and funding bodies on global program effectiveness and a determination of causation, program evaluations are fundamentally designed to assist with the planning of future programs and/or to improve pre-existing programs. The nature of many family support and other child abuse prevention programs, will preclude experimental or quasi-experimental studies of causative relationships (Jacobs 1988; Weiss 1988). Therefore, in spite of a need for 'flagship' studies, the majority of program evaluations will be modest, internally focused [fleet] studies that assess client satisfaction, document the services delivered, describe program implementation (for replication) and if possible, the immediate effects of service provision (Jacobs 1988; Weiss 1988).

There will be a variety of reasons for evaluating a program and each evaluation should be tailored to fulfil the specific purpose for which it is required and to meet the needs of the various stakeholders involved (Hutchinson 1999; Calder 1994). This then, is an argument for 'greater pluralism' in evaluation where techniques must be broad enough to enable an assessment of effectiveness across the range of available programs (Patton 1980; Jacobs 1988; Smith 1999; Swerissen 1999) and to ensure the variety of service providers are able to derive full benefits from an 'appropriate, relevant and action-linked evaluation' (O'Donoghue 1997). 'Research is a matter of asking different sorts of questions and finding different ways of answering those questions' (Smith 1999:272). The research question, and the level of explanation required, determines the methodologies and research tools used and the degree of experimental

rigour that is desired and/or possible (Brennen 1992). Such an approach has led to a general, progressive shift away from traditional experimental and quasi-experimental evaluation designs to a greater emphasis on qualitative and action research methodologies (Swerissen 1999).

It has also been contended that to make the best use of evaluation activity requires an emphasis on 'developmental sequencing', rather than the one-off evaluation of an intervention (McBride 1999). That is, evaluation is undertaken as a progressive series of analyses that build upon each other to provide a detailed analysis for service providers and other stakeholders. For example, Kirkpatrick (1967) described four levels of evaluation that could be undertaken, where the complexity of the behavioural change under investigation increased with the complexity of the evaluation strategy required. These were (from least to most complex):

- Evaluation of reaction [change in satisfaction or happiness]
- Evaluation of learning [change in knowledge or skills acquisition]
- Evaluation of behaviour [transfer of learning to the workplace]
- Evaluation of results [transfer or impact on society]

Within this model, the complexity of evaluation issues, such as the timeframe for the evaluation, the availability of reliable objective measures and the number of potentially confounding factors all increase as the complexity of the behavioural change increased.

A general starting point for developmental sequencing would be the adoption of a comprehensive evaluation model like the 'systems evaluation' model described above (Pietrzak et al. 1990), where *equal* emphasis is placed on input, process and outcome evaluation. Jacobs (1988) developed a five-tier hierarchical evaluation model that provides a comprehensive evaluation framework.

The underlying assumptions of the program evaluation model are that:

- an evaluation should be perceived as the systematic collection and analysis of program-related data for the purposes of both descriptive analysis and judging 'success';
- program evaluation should be universally adopted, and at a minimum, should incorporate a description of the beliefs underlying service provision, any changes adopted during program implementation and the services delivered;
- there are numerous legitimate purposes for evaluation – determining program impact may not always appropriate, and will depend on the stage of program development reached. An evaluation must be tailored to the program under investigation and its circumstances;
- evaluations have many legitimate audiences (that is, there are multiple stakeholders, such as staff, participants, funding bodies). Rather than attempting to conduct the

ideal evaluation that may result in admittance to 'some evaluation heaven where the souls of virtuous studies ascend' (Patton 1978:24), the goal should be to ensure that evaluation findings are used by a target audience; and

- evaluations should not distract from service delivery – the requirements of evaluation should be negotiated with agency staff and program participants.

Like Kirkpatrick's (1967) model, each level of Jacobs' model requires greater data collection and recording efforts, more precise program definitions, and a greater overall commitment to the evaluation process. Although the tiers are sequentially ordered, it is possible to engage in an evaluation at multiple levels simultaneously. If a program, or program delivery undergoes significant change, it is possible to return to the earlier phases of the evaluation model to make re-assessments. The five tiers are described below.

1. Pre-implementation

Conducted prior to the establishment of a program, this 'needs assessment' stage creates the conditions for all subsequent evaluation efforts. Here information is collected on identified social needs that justify the existence of a specific program in the target community (and the social/economic costs of not providing such a program), to delineate the basic program characteristics developed to address the identified needs, and to provide baseline data that can inform subsequent evaluations. The intention is to convince funders and the local community of the need for the proposed program.

2. Accountability

This level involves basic program monitoring via the systematic collection of data on service users and service utilisation. Specifically, it incorporates the development of an accurate description of the service provided, the characteristics of targeted participants and the systematic collection of information regarding service utilisation, the degree to which target population was successfully accessed and a basis analysis of program costs. This level of evaluation is effectively the documentation of program activities and not necessarily an attempt to investigate program (or participant) success. This tier provides enough information for programs to be minimally accountable to participants, the community and funding bodies.

As the manager of a small community family support centre struggling with the issue of program evaluation, King advocates 'bringing together a range of evaluation techniques that are not resource rich, that can be facilitated on the run and provide an adequate collage of information for planning is the challenge for the manager' (King 1998: abstract).

Jacobs (1988) notes that many grass roots agencies do not routinely engage in systematic data collection, often because they do not know what to collect, or do not have an efficient data collection system. Yet for small organisations in particular, documenting program activity – ensuring records are kept and are accurate – should be a valuable, achievable first step in evaluating their service delivery that may provide a richness of data that is relatively easily collected.

3. Program clarification

This tier involves the refinement of service delivery via formative or process evaluation processes that use previous data collection in conjunction with staff and participant feedback. The assessment of stakeholder/participant satisfaction and the provision of a venue for the beliefs, views, and feelings of the researched (Oliver 1997), 'has become a mainstream, legitimate activity that is frequently a required part of a grant from, or contract with, a purchasing authority' (Cooper 1997:13). In addition, the active involvement of multiple stakeholders (for example, staff, other agency personnel and participants) is advocated as a means of increasing evaluation validity and reliability (Nixon 1997; Stevens 1999).

Evaluation at this level is based upon assessing the extent to which service provision meets a program's articulated primary goals and objectives. The intention is to refine the program (and/or re-state the program mission, goals and objectives) in order to ensure that the needs of the local community are met effectively.

4. Progress toward objectives

Based upon the development of a comprehensive assessment of the program at previous tiers, this level represents the first focus on program effectiveness and a move towards objective measurement. This is achieved via the assessment of short term outcomes, having first derived measurable indicators of success. Where possible, these are based on standardised measures or psychometric tests; but outcome measures must first, reflect program goals rather than adapting goals to enable the use of standardised measures (see the Westinghouse Learning Corporation evaluation described above).

Thus, a combination of measurement strategies may be the best approach to adopt, where standardised measures are used together with program-specific measures designed for the evaluation (Patton 1978; Jacobs 1988; Kaufman and Zigler 1992; Clark 1997). The intent is to determine the extent to which participant progress is due to service provision, and more specifically, which elements or factors have influenced program and participant success – for whom does the program work, and why does it work?

Such an evaluation may be undertaken to develop broader community acceptance of the program, or to replicate or broaden the program (requiring additional funding). The contracting of external evaluators is common at this level, given the expertise needed to produce evidence on their effectiveness. Programs evaluated at this level are typically more established, having been in operation for some time. They are usually more experienced in service provision and have the time, interest and necessary financial support to collect and assess the required information.

5. Program impact

This tier is concerned with the development of empirical outcome studies capable of contributing to knowledge development in the field of child development, child abuse prevention or program evaluation (flagship evaluations – Weiss 1988). It is achieved via the identification of causative

program-outcome relationships through the study of short and long-term impacts on program participants.

ALTERNATIVE APPROACHES

Researchers are often criticised by service providers for providing research that is too abstract to be used (Peneul and Freeman 1997; Wadsworth 1982; 1997):

'The incomprehensibility of . . . technical research reports – pages of tables, flow charts, and scales, with their standard deviations, n's, p's, mean scores, ranges, two tailed and other mathematical tests – acts not only as an intimidatory (albeit possibly unintended) device, but also as a mechanism by which the uninitiated are prevented from adequately debating the results and conclusions' (Wadsworth 1982:236).

The adoption of a rigorous, empirical approach to social science and program evaluation research has produced a substantial reduction in the methodologies deemed appropriate for use, 'creating a serious mismatch between the richness of what could be studied and the paucity of acceptable techniques' (Jacobs 1988:41). Yet the alternative application of evaluation methods can produce clear information on social problems that is able to be used in situ [that is, 'actionable knowledge' (Argyris 1993)]. The adoption of a comprehensive evaluation framework such as that proposed by Jacobs (1988) enables the adoption of alternative approaches and methods better suited to evaluating complex social programs, like those designed to prevent child maltreatment. In the following sections a number of alternative approaches are presented that offer a means of enhancing the current state of knowledge of the effectiveness of child abuse prevention and other in situ social programs.

Enhancing the utility of empirical research

In spite of the limitations of the empirical outcome study, experimental principles will remain as an important part of the evaluation of social programs. In most cases however, rigour and a degree of experimental control will be employed to enhance quasi- or non-experimental studies in 'fleet' evaluations (Weiss 1988). Fink and McCloskey (1990) contend that a lack of uniform definitions and accurate measures of child maltreatment have hampered child abuse prevention evaluation research. Evaluators have also had to contend with a lack of a uniform standard for program success (Vimpani et al. 1996). Thus, rather than adopting an alternative approach, one option for the enhancement of current evaluation knowledge is to make improvements to evaluation outcome measurement.

Defining and measuring outcomes

First, some of the problems encountered when attempting to evaluate programs can be overcome, to some extent, by use of more objective, standardised measurements, reducing the reliance on self-report data (Ainsworth 1998). There are a vast array of measures available; these can be adapted for specific needs, although with a resultant drop in reliability and validity (Jacobs 1988). Behavioural observation of

participants by other family members, program staff, but particularly by an independent evaluator not familiar with the program (so-called 'blind' evaluation) will also reduce the effects of any self-reporting biases (Wood and Davidson 1993).

Second, it should be noted that it is generally very difficult to measure the global, community changes that may result from interventions, such as a reduction in child maltreatment rates, a drop in substance abuse or crime, because of difficulties in obtaining an accurate assessment (Tomison 1997a). Community change may take some time to become discernible via broad-based social indicators and thus, there may be a need to develop more sensitive, local community-based or problem-specific indicators in order to detect changes at any earlier stage (McGurk 1997; Tomison 1997a; Little 1999). This should involve a focus on short and mid-term outcomes (Little 1999) and may involve the utilisation of 'coalface' indicators, that is, measures developed within agencies or programs that are routinely applied in order to determine service impact on individuals or groups (McGurk 1997; Tomison 1997a).

Evidence-based practice

Given the limitations of current child abuse prevention program evaluations, evidence-based practice may offer a means of establishing a reasonable body of research upon which to base a prevention strategy (MacDonald and, Roberts 1995; Clark 1997). Although traditionally based on the 'scientist practitioner' model, with data collected in 'true experiments' (Fink and McCloskey 1990), more recent literature has made mention of a number of other tools that can be employed to inform practice (MacDonald and Roberts 1995; Clark 1997; Stevens 1999).

Under an evidenced-based approach, the object is to identify all systematic trials, published or not, including those studies which produced negative effects or a null result. Information is also generated from routine practice, making best use of qualitative data and finding 'methods of synthesizing evidence from the widest range of sources available' (Clark 1997:2). Deficits in methodological rigour are therefore compensated for, in part, by the richness and quantity of the data gathered.

Theory-based approach

As noted above, the evaluation of 'whole of community' programs has been a major area of discussion. Although traditional, rigorous empirical designs continue to be advocated by some (for example, Farrington 1997), shortcomings in the empirical evaluation of these programs has stimulated the development of new evaluation methods, such as the *theory-based approach*, which are equally useful for less complex programs. The underlying assumption of the *theory-based evaluation approach* (Connell and Kubisch 1998; Pawson and Tilley 1998) is that any empirical findings are intelligible only if the *context* of the findings and the underlying mechanisms that give rise to them are known (Pawson and Tilley 1998).

In what has been described as the current 'brand leader' in the evaluation of community-based initiatives (Pawson

and Tilley 1998), members of the US *Roundtables on Comprehensive Community Initiatives for Children and Families* (Kubisch et al. 1998) have advocated for the implementation of a 'theory of change' approach to evaluation (Connell and Kubisch 1998). This approach is defined as 'a systematic and cumulative study of the links between activities, outcomes, and contexts of the initiative' (Connell and Kubisch 1998:16), where the first step is to determine the intended outcomes of the program, the activities that will be developed to achieve the outcomes, and the contextual factors that may effect the implementation of the activities or their ability to achieve the desired outcomes (Connell and Kubisch 1998). The evaluation is thus, predicated upon the development of a clear consensual theory, the delineation of program activities and measurable short, intermediate and long-term outcomes.

An alternative theory-driven evaluation, one based on 'realist principles', is proposed by Pawson and Tilley (1998). At the core of this approach is the delineation of program context and the generative mechanisms that drive both the initiative and subsequent measures of outcome. The evaluator needs to specify how the program's operational mechanisms will produce identifiable patterns of outcome in given contextual conditions. Rather than attempt to evaluate the program in one broad-based attempt, the realist evaluator will examine the detailed theory underlying each step or phase of the initiative (this therefore needs to be clearly articulated), determining which facets can be evaluated and should be focused upon, and determining the most appropriate method to evaluate them.

As Pawson and Tilley (1998) note, neither of these two theory-based approaches necessarily means the total abrogation of quasi-experimental methods or measures. Many of the actual techniques used to evaluate aspects of CCIs may be quite methodologically rigorous. In a discussion of the evaluation of crime prevention programs, Homel and his colleagues advocate for a 'middle of the road' approach to the evaluation of CCIs (National Crime Prevention 1999).

Under their approach, theory-driven evaluations (Connell and Hubisch 1998, Pawson and Tilley 1998) provide an overarching evaluation framework and consideration is given to both community context and the generative mechanisms that lead to community change via the extensive collection of both quantitative and qualitative data. Homel et al. also propose the adoption of classic quasi-experimental methods; however, like Farrington (1997) they promote the use of comparison groups and multiple standardised outcomes, measured before and after implementation (National Crime Prevention 1999), despite the difficulties associated with applying such an analysis where CCIs are involved.

Multiple methods

Another alternative to the adoption of a classic experimental approach (that is, the use of control or comparison groups) is the use of *multiple methods* or *triangulation* – in essence, the comparison of data from multiple perspectives

(Friedman 1997; Unrau and Coleman 1997). It is not 'a tool or a strategy of validation, but an alternative to validation . . . [it] is best understood, then, as a strategy that adds rigor, breadth, and depth to any investigation' (Denzin and Lincoln 1994:2).

As the name implies, in a multiple methods approach (Campbell and Fiske 1959; Patton 1980, Hammersley and Atkinson 1983), the assumption is that confidence in findings is increased if diverse forms of data, gleaned from a variety of stakeholders and/or a number of different data collection methods, support the same conclusion (Campbell and Fiske 1959; Patton 1980; Hammersley and Atkinson 1983; Huberman and Miles 1994; Stevens 1999). The aim is for each analysis to inform the others as a means of developing a clearer, more valid perspective on program success.

Action research

Action research can be described as a collaborative venture between an evaluator and service providers where the intention is to integrate basic research, program development and evaluation into a continuous, cyclical process. Perhaps best thought of as a form of process evaluation, action research offers a flexible, interactive approach where service provision and evaluation methodology are continuously monitored, informing the refinement and modification of service delivery and the evaluative process itself (Wadsworth 1982; Rapoport 1985). 'Whole of community' approaches, where the aim is to develop social capital and to enhance community connectedness and community ownership (Tomison and Wise 1999) lend themselves to a particular form of action research: empowerment evaluation.

Empowerment evaluation

'The most important benefit of the participatory and collaborative model of evaluation is that people make changes based upon the findings of an evaluation when they have been involved in generating them' (Cooper 1997:13).

Empowerment evaluation has been defined by Fetterman (1994) as using the process of evaluation to assist people and communities to achieve self-determination and to gain control of their futures. The method involves the development of partnerships between an external evaluator and participants, with the latter taking an active role in developing and informing the evaluation, but also being trained in evaluation skills. Thus, the external evaluator acts as a facilitator rather than as a reviewer; as an advocate for disadvantaged groups, assisting them to take control of their situation and to make the most of available opportunities and resources (Fetterman 1994). Empowerment evaluation may lead to: increased staff cooperation with evaluations; increased staff understanding of evaluation reports; an increased likelihood that recommendations will be implemented and that future evaluation efforts will be undertaken (Fetterman 1994).

Empowerment evaluation is about 'doing with' rather than 'doing to'. Programs run in indigenous communities

often appear to have adopted such methods. The historical practice of evaluators entering an Australian Aboriginal or Torres Strait community, often without adequate consultation, studying indigenous people as the 'objects' of research, and subsequently disappearing never to be heard from again, has left many Aboriginal and Torres Strait communities feeling hurt, mistrustful of the research process and unwilling to subject themselves to it again (National Aboriginal Health Strategy Working Party 1989). Empowerment evaluation may offer a means of improving relationships with indigenous peoples and a vehicle to re-build trust.

Such an approach however, is often difficult to achieve and may not lead to any real change in the participant 'community' as a function of the significant energy tied up in the process of developing a participatory evaluation model. There is also a tension between the need to produce empirical evidence of outcomes, perhaps more the focus of the theory-driven evaluation approach, and an empowerment evaluation approach which is more about the *process* of evaluation research. Empowerment evaluation has been perceived as being an anti-scientific evaluation movement or philosophy, but not as an evaluation method per se (Sechrest 1997). Although supportive of attempts to teach service providers program evaluation skills, empowerment evaluation is seen by some as merely an approach to teaching evaluation skills, not a form of evaluation in itself (Scriven 1999).

RESOURCES

As noted above, although it may be appropriate to expect that a rigorous evaluation is conducted by large, complex organisations, the level of resourcing, knowledge and expertise available to the majority of grass roots service providers has often precluded attempts to produce a cohesive, detailed program evaluation (Weiss and Hite 1986; O'Donoghue 1997). O'Donoghue notes that:

'Doing evaluation well is an exacting task. It requires skills, knowledge, perspective – and time and money. So it becomes very easy to see evaluation as a job for experts. In fact, if you can organise a tame expert, it seems like the best of available arrangements – you can have a clear conscience that you are dutifully evaluating your performance from first principles to practical detail; it is all being done according to professional standards; and with any luck you won't have to do much about the awkward results because you will be charging forward vigorously to the next objective before the report is in.' (O'Donoghue 1997: Supplementary papers)

It is often recommended in the evaluation literature that creative partnerships be developed between researchers, workers and program developers as a means of producing quality evaluations that can subsequently better inform policy and practice (Tomison 1997b; Chalk and King 1998). Similarly, researchers and service providers alike, have identified a need for specialist program evaluation advice and information to be made readily available to service providers attempting to develop a program evaluation.

In Australia there are a variety of government units and university-based centres specialising in conducting research and advising service providers on aspects of program delivery and evaluation. For example, the National Centre for Health Program Evaluation, Fairfield Hospital, the Victorian Parenting Centre, Centres for Health Program Evaluation, (Melbourne and Monash Universities). In order to partially fulfil the needs of those evaluating child abuse prevention programs, the National Child Protection Clearinghouse has recently developed an evaluation consultancy service that provides free assistance to service providers attempting to design an evaluation. The service includes assisting them in:

- creating partnerships with researchers and/or evaluation consultants capable of conducting the evaluation;
- developing a program evaluation manual for service providers;
- program evaluation workshops, designed to provide participants with an overview of the evaluation process and a venue for the discussion of the design and implementation of program evaluations;
- discussion of program evaluation issues on the Clearinghouse email discussion list (*childprotect@aifs.org.au*); and
- developing a special program evaluation reference collection as part of the overall Clearinghouse library.

Another useful service advocated for agencies planning an evaluation is the development of a foundation or oversight agency to fund evaluations in a particular area, or field of endeavour, as a productive and cost-effective way to upgrade performance. Scriven (1999) for example, notes the value of government 'inspector-general' type agencies that can hire evaluators to complete evaluations, but are also able to independently judge the resultant project on its merits, given there is no link with funding bodies.

Stakeholder participation

The benefits of a partnership between service providers and external evaluator are not unilateral, however. The establishment of a positive relationship with service providers, and a level of agency support for the evaluation process, would seem to be a paramount consideration if an external evaluator is to fully understand the realities of program and service provision and effect a quality research result (Halbrook et al. 1997; Rist 1997). 'An outsider (external evaluator) can only effect useful change *through* and *with* insiders' [original author's italics] (Wadsworth 1997:19). Such a partnership is essentially mandatory in order to ensure that the perceptions of stakeholders and participants are considered as part of the evaluation (Cooper 1997; Oliver 1997; Nixon 1997; Stevens 1999).

The participation of children and young people

The assessment of participant satisfaction and the provision of a venue for the beliefs, views, and feelings of the researched has become a key component of program evaluation research (Cooper 1997; Oliver 1997), partly as a

means of increasing an evaluation's validity and reliability (Nixon 1997; Stevens 1999). As was also mentioned above, where possible it is important to incorporate the responses of children and young people who either participate in, or who are directly affected by, a particular program.

Evaluations frequently incorporate measures of child development, functioning or wellbeing, but it is only relatively recently that researchers have recognised the necessity and importance of utilising evaluation methods that are 'child friendly' and that provide children and young people with an opportunity to be heard and to actively contribute to the program evaluation process (O'Brien 1997; Tomison 1997a; Smith and Taylor 1998).

CONCLUSION

What then, are the messages to be taken away from this paper? First, it is clear that in spite of the general acceptance that the evaluation of programs should be a core element of service provision, and that rigorous, scientific evaluation is essential for the determination of causal relationship, the difficulties associated with applying such methods in situ, have meant that very few 'gold standard' evaluations have been done in Australia or internationally (Fink and McCloskey 1990; Harrington and Dubowitz 1993; James 1994; Melton and Flood 1994; Tomison 1995; 1997a; 1997b; 1998; Chalk and King 1998).

With the exception of a set of 'two dozen or so flagship research and demonstration programs' (Weiss 1988), most child abuse prevention program evaluations that are undertaken are non-experimental studies which provide no firm basis for determining the relative impact of a program or intervention, or of examining the impact of a particular program or activity on specific populations (Chalk and King 1998). As a result, although policymakers' expectations of what may be accomplished by evaluation continue to grow, evaluators (and service providers) have generally become much more conservative in their claims about the potential of evaluations to demonstrate 'success', the incremental nature of evaluation evidence and how it should be used by policy makers (Rist 1997; Little 1999). Therefore, experimental rigour in isolation, should no longer be perceived as driving the majority of program evaluations. Rather, methodological rigour is better thought of as a key component of comprehensive, systematic evaluation approaches, where equal attention is paid to input, process and output assessments.

Second, it is clear that the majority of evaluations will be modest, internally focused [fleet] studies that assess client satisfaction, document the services delivered, describe program implementation (for replication) and if possible, the immediate effects of service provision (Jacobs 1988; Weiss 1988). In order to cater for the needs of service providers undertaking such evaluations there must be 'greater pluralism' in evaluation, where the research questions and the level of explanation required will determine which of a variety of methods and tools will be used, along with the degree of experimental rigour that is desired and/or is possible (Brennen 1992).

Such an approach will continue to drive a general, progressive shift away from traditional experimental and quasi-experimental evaluation designs to a greater emphasis on alternative methodologies (Swierissen 1999) and an emphasis on 'developmental sequencing'. Here, rather than undertaking a one-off evaluation of an intervention (McBride 1999), evaluation is undertaken as a progressive series of analyses that build upon each other to provide a detailed analysis for service providers and other stakeholders (similar to the continuous improvement concept in quality assurance projects). Thus, the emphasis is on the developmental nature of program evaluation, where simple assessments of program structure and content that can be completed by all service providers, form the basis for potentially more rigorous assessment of short and long-term program effects.

The third point to note therefore, is the importance of adequately grounding evaluation by ensuring the adequate documentation of program activity. Ensuring records are kept and are accurate is a valuable, achievable, first step in evaluating service delivery that may provide a richness of data able to be collected relatively easily. Evaluations may be highly complex, experimentally rigorous projects or simple, informal internal assessments, but regardless of the size, cost and degree of rigour, all evaluations should follow the same general guidelines or rules to ensure that as accurate and precise an evaluation outcome is produced (Schalock and Thornton 1988).

In preparing for an evaluation the following aspects of the program therefore should be clearly documented (Schalock and Thornton 1988; McCurdy 1995; US Department of Health and Human Services 1995; Chalk and King 1998):

- *knowledge of other services* that are currently operating in the community and the *evaluation context*;
- *the theory of change that guides such interventions*;
- *the problem to be addressed*;
- *a clear description of the program* – this should include the assumptions used to develop the target population and subsequently, the program or activity;
- *the stages of implementation*;
- *the client referral, screening, and baseline assessment processes*;
- *the client population* – the characteristics of the participant population should be mapped as a diversity of clients within a program will require thought as to how the evaluation may take this diversity into account during analysis;
- *a statement of measurable objectives* – in process terms, this will involve the identification of the interventions to be carried out with the target population, the staff assigned to do it and who and how many participants will be accessed: in outcome terms, the service provider will outline the specific changes in knowledge, attitudes or behaviour that are expected to occur.

Overall, there has been increased acceptance of the importance of program evaluation activity. More recently, there

has been some recognition that while the 'gold standard' of empirical research may be the best opportunity of identifying causative relationships, for the majority of programs evaluation efforts will be focused on specific program refinement and/or an assessment of short-term outcomes.

To promote evaluation best practice in these circumstances requires acknowledgment of the circumstances of the average service provider – the availability of physical and professional resources; consideration of the appropriateness and limitations of empirical studies; a willingness to explore the variety of evaluation methods and techniques currently available in order to find those best suited for the purposes of the evaluation; and an understanding of the benefits that may be attained via the developmental sequencing of program evaluations.

Notes

1. Empirical evaluation is defined in this paper as involving the conduct of 'true experiment' (Fink & McCloskey 1990) involving pre- and post-test models, and the adoption of matched control and experimental samples. Under such an approach the overall intention would be to evaluate with large sample sizes over time, enhancing the potential for future replication.

2. Child abuse prevention is commonly classified into three main levels under a 'public health' model: primary, secondary and tertiary prevention (Helfer 1982). *Primary prevention* is targeted at the community as a whole; the aim of primary prevention programs is to stop abuse or neglect before it starts (Calvert 1993). Primary programs generally comprise community education campaigns aimed at both children and adults, or school-based personal safety and healthy relationships programs for children. *Secondary prevention* programs target specific 'at risk' sections of the population. That is, those with special needs or who are in need of greater support, such as young parents, single parents, people with physical or intellectual disabilities. Secondary prevention programs can be categorised as enhancing family functioning by providing various forms of family support and, in particular, by teaching parenting skills and increasing parents' knowledge of child development and behavioural expectations (Calvert 1993). *Tertiary prevention* refers to prevention initiatives that are aimed at preventing the recurrence of abuse in those families where children have already been maltreated (Calvert 1993).

3. It should be noted that detecting changes in the prevalence of child maltreatment in local communities is difficult, given that no accurate measure of actual child maltreatment exists. Child protection statistics or other agency statistics are available, but these are an indirect underestimate at best.

References

- Ainsworth, F. (1998), 'Program evaluation for child and family services: What can be done?', *Children Australia*, vol. 23, no. 2, pp. 39-43.
- Argyris, C. (1993), *Knowledge for Action: A Guide to Overcoming Barriers to Organizational Change*, Jossey-Bass, San Francisco.
- Bickman, L. (1992), 'Designing outcome evaluations for children's mental health services: Improved internal validity', *New Directions for Program Evaluation*, vol. 54, pp. 57-68.
- Bickman, L. (1999), 'Practice makes perfect and other myths about mental health services', *American Psychologist*, vol. 54, pp. 965-978.
- Bickman, L., Summerfeldt, W.T., Firth, J.M. & Douglas, S.M. (1997), 'The Stark County Evaluation Project: Baseline results of a randomized experiment', in C.T. Nixon and D.A. Northrup (eds), *Evaluating Mental Health Services: How do programs for Children*

'Work' in the Real World?, Children's Mental Health Series, Volume 3, Sage, Thousand Oaks.

Bowes, J.M. & Hayes, A. (1999), 'Contexts and consequences: impacts on children, families and communities', in J.M. Bowes & A. Hayes (eds), *Children, Families, and Communities: Contexts and Consequences*, Oxford University Press, South Melbourne.

Brennen, J. (1992), 'Combining qualitative and quantitative approaches: an overview', in J. Brennen (ed.), *Mixing Methods, Qualitative and Quantitative Research*, Avebury Ashgate Publishing, Aldershot.

Bronfenbrenner, U. (1979), *The Ecology of Human Development*, Harvard University Press, Cambridge, MA.

Butler, I. and Williamson, H. (1994), *Children Speak: Children, Trauma and Social Work*, Longman Group, Harlow, Essex.

Calder, J. (1994), *Programme Evaluation and Quality*, Kogan Page, London.

Calvert, G. (1993), *Preventing Child Abuse: A National Strategy*, National Child Protection Council, Canberra.

Campbell, D.T. & Fiske, D.W. (1959), 'Convergent and discriminant validation by the multitrait-multimethod matrix', *Psychological Bulletin*, vol. 56, no. 2, pp. 297-311.

Campbell, D.T. & Stanley, J.C. (1966), *Experimental and Quasi-Experimental Designs for Research*, (2nd edn), Rand McNally, Chicago, IL.

Chalk, R. & King, P.A. (Eds.) (1998), *Violence in families: Assessing prevention and treatment programs*. National Academy Press, Washington DC.

Clark, R. (1997), 'A research agenda – what does it mean?', Paper presented to Research Agenda Workshop, Youth & Family Services Division, Department of Human Services Victoria, Melbourne, 20 March 1997.

Connell, J.P. & Kubisch, A.C. (1998), 'Applying a Theory of Change approach to the evaluation of initiatives: progress, prospects, and problems', in K. Fullbright-Anderson, A.C. Kubisch & J.P. Connell (eds), *New Approaches to Evaluating Community Initiatives: Volume 2: Comprehensive Community Initiatives: Theory, Measurement and Analysis*, Roundtable on Comprehensive Community Initiatives for Children and Families, The Aspen Institute, Washington DC.

Cooper, L. (1997), 'Evaluation and the third sector', Keynote speech presented to the Australasian Evaluation Society 1997 International Conference, Adelaide, South Australia, Australasian Evaluation Society, Curtin, ACT.

Crouch, R.A. (1998), 'The progress of experiment: science and therapeutic reform in the United States, 1900-1990', Book review, *British Medical Journal*, vol. 320, no. 7137, p. 1101.

Curtis, S. (1997), 'Action speaks louder than words in the third sector: measures of effective practice that really count', in *Australasian Evaluation Society International Conference 1997: Proceedings*, pp. 196-202, Australasian Evaluation Society, Curtin, ACT.

Dallaire, N., Chamberland, C., Cameron, S. & Hébert, J. (1995), 'Social prevention: a study of projects in an urban environment', in J. Hudson & B. Galaway (eds), *Child Welfare in Canada: Research and Policy Implications*, Thompson Educational Publishing, Toronto.

Denzin, N.K. & Lincoln, Y.S. (1994), 'Introduction: entering the fields of qualitative research', in N.K. Denzin & Y.S. Lincoln (eds), *Handbook of Qualitative Research*, Sage, Thousand Oaks, CA.

Devine, J.A., Wright, J.D. & Joyner, L.M. (1994), 'Issues in implementing a randomised experiment in a field setting', in K.J. Conrad (ed.), *Critically Evaluating the Role of Experiments*, Jossey Bass, San Francisco.

Ellwood, A. (1988), 'Prove to me that MELD makes a difference', in H.B. Weiss & F.H. Jacobs (eds), *Evaluating Family Programs*, de Gruyter, New York.

Farmer, E.M.Z., Burns, B.J., Guiles, H.B., Behar, L. & Gerber, D.I. (1997), 'Conducting randomized clinical trials in children's mental health: experiences and lessons from one venture', in C.T. Nixon & D.A. Northrup (eds), *Evaluating Mental Health Services: How do Programs for Children 'Work' in the Real World?*, Children's Mental Health Series, Volume 3, Sage, Thousand Oaks.

Farrington, D.P. (1997), 'Evaluating a community crime prevention program', *Evaluation*, vol. 3, no. 2, pp. 157-173.

Fetterman, D.M. (1994), 'Empowerment evaluation', *Evaluation Practice*, vol. 15, no. 1, pp. 1-15.

Fink, A. & McCloskey, L. (1990), 'Moving child abuse and neglect prevention programs forward: improving program evaluations', *Child Abuse and Neglect*, vol. 14, pp. 187-206.

Firestone, W.A. & Herriott, R.E. (1983), 'The formalization of qualitative research: an adaptation of "soft" science to the policy world', *Evaluation Review*, vol. 7, no. 4, pp. 437-466.

Fraser, M.W., Nelson, K.E. & Rivard, J.C. (1997), 'Effectiveness of family preservation services', *Social Work Research*, vol. 21, no. 3, pp. 138-153.

Friedman, R.M. (1997), 'Services and service delivery systems for children with serious emotional disorders: Issues in assessing effectiveness', in C.T. Nixon & D.A. Northrup (eds), *Evaluating Mental Health Services: How do Programs for Children 'Work' in the Real World?*, Children's Mental Health Series, Volume 3, Sage, Thousand Oaks.

Frost, N. (1989), *The Politics of Child Welfare: Inequality, Power and Change*, Harvester Wheatsheaf, London.

Godlee, F. (1998), 'Getting evidence into practice', Editorial, *British Medical Journal*, vol. 317, p. 6.

Guba, E.G. & Lincoln, Y.S. (1981), *Effective Evaluation*, Jossey-Bass, San Francisco, CA.

Halbrook, B., Sheley, J.F., Wycoff, S., Restaino, D.P. et al. (1997), *Expert Advisory Report: A Feasibility Study of the Transition to Outcomes-Based Evaluation for OCJP Child Abuse Programs funded by the Victims of Crime Act: Child Abuse*, Center for Collaboration for Children, California State University, Fullerton, CA.

Hammersley, M. & Atkinson, P. (1983), *Ethnography: Principles and Practice*, Tavistock, New York.

Hanson, R.K. (1997), 'How to know what works with sexual offenders', *Sexual Abuse*, vol. 9, no. 2, pp. 129-147.

Harrington, D. & Dubowitz, H. (1993), 'What can be done to prevent child maltreatment?', in R. L. Hampton (ed.), *Family Violence: Prevention and Treatment Issues in Children's and Families Lives*, vol. 1, Sage Publications, Newbury Park, CA.

Hawkins, P. & Leigh, J. (1997), 'Challenges facing government employed evaluators of community based programmes in New Zealand', in *Australasian Evaluation Society 1997 International Conference – 'Evaluation: Equipping communities and government'*, 1-3 October, Adelaide, Published in Australasian Evaluation Society, Supplementary Papers to Proceedings (pp. 1-4).

Heflinger, C.A. (1987), 'Psychotherapy, research ethics: continuing the debate on controlled clinical trials', *American Psychologist*, vol. 42, no. 10, p. 957.

Helfer, R. E. (1982), 'A review of the literature on the prevention of child abuse and neglect', *Child Abuse and Neglect*, vol. 6, pp. 251-61.

Hoagwood, K. (1994), Introduction to the special section: issues in designing and implementing studies in non-mental health care sectors, *Journal of Clinical Child Psychology*, vol. 23, pp. 114-120.

Hoagwood, K., Hibbs, E., Brent, D. & Jensen, P. (1995), 'Introduction to the Special Section: efficacy and effectiveness in studies of child and adolescent psychotherapy', *Journal of Consulting and Clinical Psychology*, vol. 63, pp. 683-687.

Hobbs, N., Dokecki, P., Hoover-Dempsey, K., Moroney, R., Shayne, M. & Weeks, K. (1984), *Strengthening Families*, Jossey-Bass, San Francisco, CA.

- Hollister, R.G. & Hill, J. (1995), 'Problems in the evaluation of community-wide initiatives', in A.C. Kubisch, L.B. Schorr & C.H. Weiss (eds), *New Approaches to Evaluating Community Initiatives*, The Aspen Institute, New York.
- Huberman, A.M. & Miles, M.B. (1994), 'Data management and analysis methods', in N.K. Denzin & Y.S. Lincoln (eds), *Handbook of Qualitative Research*, Sage, Thousand Oaks, CA.
- Hutchinson, L. (1999), 'Evaluating and researching the effectiveness of educational interventions', *British Medical Journal*, vol. 318, pp. 1267-1269.
- Jacobs, F.H. (1988), 'The five-tiered approach to evaluation: context and implementation', in H.B. Weiss & F.H. Jacobs (eds.), *Evaluating Family Programs*, Modern Applications of Social Work series, Aldine de Gruyter, New York.
- James, M. (1994), 'Analysis of the programs database', *Child Abuse Prevention*, vol. 2, no. 1, pp. 1-3.
- Kahn, A. & Kamerman, S. (eds) (1996), *Children and Their Families in Big Cities: Strategies for Service Reform*, Cross National Studies Research Program, Columbia University, New York.
- Kaufman, J. & Zigler, E. (1992), 'The prevention of child maltreatment: programming, research and policy', in D. J. Willis, E. W. Holden & M. S. Rosenberg (eds), *Prevention of Child Maltreatment: Developmental and Ecological Perspectives*, John Wiley & Sons, New York.
- Kempe, R. S., Silverman, F. N., Steele, B. F., Droegemuller, W. & Silver, H. K. (1962), 'The battered child syndrome', *Journal of the American Medical Association*, vol. 18, no.1, pp.17-24.
- King, B. (1998), 'Making evaluation work', paper presented to the 6th Australian Institute of Family Studies Conference 'Changing families, challenging futures', Melbourne, November.
- Kirkpatrick, D.I. (1967), 'Evaluation of training', in R. Craig & I. Bittel (eds), *Training and Development Handbook*, MacGraw-Hill, New York.
- Kubisch, A.C., Fullbright-Anderson, K. & Connell, J.P. (1998), 'Evaluating community initiatives: a progress report', in K. Fullbright-Anderson, A.C. Kubisch & J.P. Connell (eds), *New Approaches to Evaluating Community Initiatives: Volume 2: Theory, Measurement and Analysis*, Roundtable on Comprehensive Community Initiatives for Children and Families, Aspen Institute, Washington DC.
- Little, M. (1999), 'Prevention and early intervention with children in need: definitions, principles and examples of good practice', *Children and Society*, vol. 13, pp. 304-316.
- MacDonald, G. & Roberts, H. (1995), *What works in the early years? Effective interventions for children and their families in health, social welfare, education and child protection*. Barnardos, Essex.
- McBride, N. (1999), 'Health promotion evaluation and research in schools: issues for consideration', *Australian Journal of Primary Health – Interchange*, vol. 5, no. 4, pp. 37-45.
- McCurdy, K. (1995), *Summary of Home Visiting Research*, National Committee to Prevent Child Abuse, Chicago.
- McGurk, H. (1997), 'Context for a research program for the Division of Youth and Family Services, Department of Human Services Victoria', Paper presented to Research Agenda Workshop, Youth & Family Services Division, Department of Human Services Victoria, Melbourne, 20 March.
- McKee, M., Britton, A., Black, N., McPherson, K, Sanderson, C. & Bain, C. (1999), 'Interpreting the evidence: choosing between randomised and non-randomised studies', *British Medical Journal*, vol. 319, pp. 312-315.
- Melton, G.B. & Flood, M.F. (1994), 'Research policy and child maltreatment: developing the scientific foundation for effective protection of children', *Child Abuse & Neglect*, vol. 18, suppl. 1, pp. 1-28.
- Midford, R. & Boots, K. (1999), 'COMPARI: Insights from a three-year community based alcohol harm reduction project', *Australian Journal of Primary Health – Interchange*, vol. 5, no. 4, pp. 46-58.
- Miner, M.H. (1997), 'How can we conduct treatment outcome research?', *Sexual Abuse*, vol. 9, no. 2, pp. 95-110.
- Mishler, E.G. (1979), 'Meaning in context: Is there any other kind?', *Harvard Educational Review*, vol. 49, pp. 1-19.
- National Aboriginal Health Strategy Working Party, (1989), *A National Aboriginal Health Strategy*, AGPS, Canberra.
- National Crime Prevention (1999), *Pathways to Prevention: Developmental and Early Intervention Approaches to Crime in Australia*, Commonwealth Attorney-General's Department, Canberra.
- Nixon, C.T. (1997), 'How does evaluation of mental health services for children work in the real world?', in C.T. Nixon & D.A. Northrup (eds), *Evaluating Mental Health Services: How do Programs for Children 'Work' in the Real World?*, Children's Mental Health Series, Volume 3, Sage, Thousand Oaks.
- Nocon, A. & Qureshi, H. (1996), *Outcomes of Community Care for Users and Carers*, Open University Press, Buckingham.
- O'Brien, A. (1997), 'Consumer participation for young people in care', *Family Matters*, no. 46, Autumn, pp. 56-58.
- Ochiltree, G. (1999), 'Lessons from Head Start in the USA', *Brotherhood Comment*, May, pp. 10-11.
- O'Donoghue, L. (1997), 'Evaluation: equipping communities and government', Opening address in *Australasian Evaluation Society 1997 International Conference 'Evaluation: Equipping communities and government'*, 1-3 October, Adelaide. Published in *Australasian Evaluation Society, Supplementary Papers to Proceedings*.
- Olds, D.L., Eckenrode, J., Henderson, C.R., Kitzman, H., Powers, J., Cole, R., Sidora, K., Morris, P., Pettitt, L.M. & Luckey, D. (1997), 'Long-term effects of home visitation on maternal life course and child abuse and neglect', *Journal of the American Medical Association*, vol. 278, no. 8, pp. 637-643.
- Olds, D.L., Henderson, C.R., Chamberlin, R. & Tatelbaum, R. (1986a), 'Preventing child abuse and neglect: a randomised trial of nurse intervention', *Pediatrics*, no. 78, pp. 65-78.
- Olds, D.L., Henderson, C.R., Tatelbaum, R. & Chamberlin, R. (1986b), 'Improving the delivery of prenatal care and outcomes of pregnancy: a randomised trial of nurse home visitation', *Padiatrics*, no. 77, pp. 16-28.
- Oliver, P. (1997), 'Giving "voice" to the researched in evaluation research', Paper presented to the *Australasian Evaluation Society 1997 International Conference, Adelaide, South Australia*, Australasian Evaluation Society, ACT.
- Parker, R., Ward, H., Jackson, S., Aldgate, J. & Wedge, P. (1991), *Assessing Outcomes in Child Care*, HMSO, London.
- Patton, M.Q. (1978), *Utilization focused evaluation*, Sage, Beverley Hills.
- Patton, M.Q. (1980), *Qualitative evaluation methods*, Sage, Beverley Hills.
- Pawson, R. & Tilley, N. (1994), 'What works in evaluation research?', *British Journal of Criminology*, vol. 34, no. 3, pp. 291-306.
- Pawson, R. & Tilley, N. (1998), 'Caring communities, paradigm polemics, design debates', *Evaluation*, vol. 4, no. 1, pp. 79-90.
- Peneul, W.R. & Freeman, T. (1997), 'Participatory action research in youth programming: a theory in use', *Child and Youth Care Forum*, vol. 26, no. 3, pp. 175-185.
- Pietrzak, J., Ramler, M., Renner, T., Ford, L. & Gilbert, N. (1990), *Practical Program Evaluation: Examples from Child Abuse Prevention*, Sage Sourcebooks for the Human Services series 9, Sage, Newbury Park, CA.
- Rapoport, R. (1985), 'Research and action', in R. Rapoport (ed.), *Research and Action: Innovation for Children, Youth, and Families*, Cambridge Press, London.
- Rees, S. (1994), 'Economic rationalism: an ideology of exclusion', *Australian Journal of Social Issues*, vol. 29, no. 2, pp. 171-85.

Reppucci, N.D., Woolard, J.L. & Fried, C.S. (1999), 'Social, community, and preventive interventions', *Annual Review of Psychology*, Vol. 50, pp. 387-418.

Reynolds, A.J. (1998), 'Developing early childhood programs for children and families at risk: research-based principles to promote long-term effectiveness', *Children and Youth Services Review*, vol. 20, no. 6, pp. 503-523.

Rist, R.C. (1997), 'Evaluation and organizational learning: some international observations', Paper presented to the *Australasian Evaluation Society 1997 International Conference, Adelaide, South Australia*, Australasian Evaluation Society, ACT.

Schalock, R.L. & Thornton, C.V.D. (1988), *Program Evaluation: A Field Guide for Administrators*, Plenum Press, New York.

Scriven, M. (1999), 'The fine line between evaluation and explanation', *Research on Social Work Practice*, vol. 9, no. 4, pp. 521-524.

Sechrest, L. (1997), 'Book review of Fetterman et al. 1996', *Environment and Behaviour*, vol. 29, no. 3, pp. 422-426.

Shils, E. (1980), *The calling of Sociology*, University of Chicago Press, Chicago, IL.

Smith, A.B. & Taylor, N.J. (1998), 'A new discourse for child advocacy, children's services and research', Paper presented to the *6th Australian Institute of Family Studies Conference, Changing families, challenging futures*, 25-27 November, Melbourne.

Smith, T. (1999), 'Neighbourhood and preventive strategies with children and families: What works?', *Children and Society*, vol. 13, pp. 265-277.

Stevens, M. (1999), 'Assessing outcomes in child welfare: a critical review', *Research, Policy and Planning*, vol. 17, no. 1, pp. 26-32.

Swerissen, H. (1999), 'Health promotion evaluation' Guest editorial, *Australian Journal of Primary Health – Interchange*, vol. 5, no. 4, pp. 6-8.

Tomison, A. M. (1995), 'The current state of child abuse prevention programs in Australia: an analysis of the National Child Protection Clearinghouse databases', Paper presented to the *Fifth Australasian Conference on Child Abuse and Neglect*, Melbourne, 17-19 October.

Tomison, A.M. (1997a), *Overcoming Structural Barriers to the Prevention of Child Abuse and Neglect: A Discussion Paper*, NSW Child Protection Council, Sydney.

Tomison, A.M. (1997b), *Preventing Child Abuse and Neglect in NSW: Findings from a State Audit*, NSW Child Protection Council, Sydney.

Tomison, A.M. (1998), *Valuing Parent Education: A Cornerstone of Child Abuse Prevention*, NCPCH Issues Paper, No. 10, Australian Institute of Family Studies, Melbourne.

Tomison, A.M. & Wise, S. (1999), *Community-based Approaches in Preventing Child Maltreatment*. NCPCH Issues Paper No. 11, Australian Institute of Family Studies, Melbourne.

Unrau, Y.A. & Coleman, H. (1997), 'Qualitative data analysis', in R.M. Grinnell Jr., *Social Work Research and Evaluation: Quantitative and Qualitative Approaches* (5th edn), F.E. Peacock, Ithaca, Ill.

US Department of Health and Human Services (1995), *The Program Manager's Guide to Evaluation: An Evaluation Handbook Series from the Administration on Children, Youth and Families*, US Department of Health and Human Services, Washington DC.

Vimpani, G., Frederico, M., Barclay, L. & Davis, C. (1996), *An Audit of Home Visitor Programs and the Development of an Evaluation Framework*, AGPS, Canberra.

Wadsworth, Y. (1982), 'The politics of social research: a social research strategy for the community health, education and welfare movement', *Australian Journal of Social Issues*, vol. 17, no. 3, pp. 232-246.

Wadsworth, Y. (1997), *Everyday Evaluation on the Run* (2nd edn), Allen & Unwin, St Leonards, NSW.

Weiss, C.H. (1983), 'The stakeholder approach to evaluation: origins and promise', in A.S. Bryk (ed.), *Stakeholder-based Evaluation* (New Directions for Program Evaluation, no. 17, pp. 3-14), Jossey-Bass, San Francisco.

Weiss, H.B. (1988), 'Family support and education programs: working through ecological theories of human development', in H.B. Weiss & F.H. Jacobs (eds.), *Evaluating Family Programs*, Modern Applications of Social Work series, Aldine de Gruyter, New York.

Weiss, H.B. & Hite, S. (1986), 'Evaluation: Who's doing it and how?', *Family Resource Coalition Report*, vol. 5, no. 3, pp. 4-7.

Whipple, E.E. & Wilson, S.R. (1996), 'Evaluation of a parent education and support program for families at risk of physical child abuse', *Families in Society*, no. 77, April, pp. 227-39.

Willis, D. J., Holden, E. W. and Rosenberg, M. (1992), 'Child maltreatment prevention: introduction and historical overview', in D. J. Willis, E. W. Holden and M. Rosenberg (eds), *Prevention of Child Maltreatment: Developmental and Ecological Perspectives*, John Wiley & Sons, New York.

Wood, C. & Davidson, J. (1993), 'Conflict resolution in the family: a PET evaluation study', *Australian Psychologist*, vol. 28, no. 2, pp. 100-104.

Zigler, E.F. & Styfco, S. (1996), 'Head Start and Early Childhood Intervention: the changing course of social science and social policy', in E.F. Zigler, S.L. Kagan & N.W. Hall (eds), *Children Families and Government: Preparing for the Twenty-first Century*, Cambridge University Press, New York.

ADAM M. TOMISON is the Research Advisor for the National Child Protection Clearinghouse, and a Senior Research Fellow at the Australian Institute of Family Studies.

BECOME PART OF THE CHILD ABUSE PREVENTION NETWORK!

The National Child Protection Clearinghouse at the Australian Institute of Family Studies serves as an interchange point for information, research and initiatives in the child abuse prevention field. It collects and distributes information, and aims for a two-way involvement with the community concerned with child protection.



To participate in the work of the Clearinghouse –

- send us materials relevant to child abuse prevention;
- complete and return a questionnaire on research and program activities relevant to child protection with which you are involved;
- join the National Clearinghouse mailing list – you will receive two newsletters and two issues papers free of charge each year.



National Clearing House Publications

The National Child Protection Clearinghouse at the Australian Institute of Family Studies produces three types of publication – Issues Papers, Discussion Papers, and Newsletters. These publications are available from the Australian Institute of Family Studies free of charge.

Child Abuse Prevention – Issues Papers

- No 12. *Evaluating child abuse prevention programs*, by Adam M. Tomison, Autumn 2000.
- No 11. *Community-based approaches in preventing child maltreatment*, by Adam M. Tomison and Sarah Wise, Autumn 1999.
- No 10. *Valuing parent education: a cornerstone of child abuse prevention*, by Adam M. Tomison, Spring 1998.
- No 9. *Long-term effects of child sexual abuse*, by Paul E. Mullen and Jillian Fleming, Autumn 1998.
- No 8. *Emotional abuse: the hidden form of maltreatment*, by Adam M. Tomison and Joe Tucci, Spring 1997.
- No 7. *Child maltreatment and disability*, by Adam M. Tomison, Spring 1996.
- No 6. *The intergenerational transmission of child maltreatment*, by Adam M. Tomison, Autumn 1996.
- No 5. *Update on child sexual abuse*, by Adam M. Tomison, Spring 1995.
- No 4. *Spotlight on child neglect*, by Adam M. Tomison, Autumn 1995.
- No 3. *Child abuse prevention: a perspective on parent enhancement programs from the United States*, by Marianne James, 1994.
- No 2. *Domestic violence as a form of child abuse: identification and prevention*, by Marianne James, 1994.
- No 1. *Child abuse and neglect: incidence and prevention*, by Marianne James, 1994.

Child Abuse Prevention – Discussion Papers

- No 3. *Child maltreatment and family structure*, by Adam M. Tomison, 1996.
- No 2. *Child maltreatment and substance abuse*, by Adam M. Tomison, 1996.
- No 1. *Child maltreatment and mental disorder*, Adam M. Tomison, 1996.



Child Abuse Prevention – Newsletters

The National Child Protection Clearinghouse Newsletter, *Child Abuse Prevention*, is produced twice a year to keep members up-to-date with new information and provide a forum for ideas.

Clearinghouse publications can be obtained free of charge by contacting the Australian Institute of Family Studies. Phone: (03) 9214 7888.

APPLICATION FORM

NATIONAL CHILD PROTECTION CLEARINGHOUSE

- ☐ Please send questionnaire concerning my research. ☐ Please add my name to your mailing list
- ☐ Please send questionnaire concerning my program activities.

Title _____ Full name _____

Organisation _____

Address _____ (Postcode) _____

Phone _____ Fax _____



Send your application to:

National Child Protection Clearinghouse,

Australian Institute of Family Studies, 300 Queen Street, Melbourne, Victoria 3000 Australia.

Phone: (03) 9214 7888. Fax: (03) 9214 7839. Internet: www.aifs.org.au/