RANDOMIZED AND QUAS EXPERIMENTAL E VALUATIONS OF PROGRAM IMPACT IN CHILD WELFARE IN CANADA: A REVIEW

Robert J. Flynn
Danielle Bouchard
University of Ottawa
Ottawa, Ontario

Abstract: Respondents to a recent survey identified the evaluation of service effectiveness as the most pressing priority for child welfare research in Canada. After a comprehensive search, we located and reviewed 10 peer-reviewed impact evaluations, published during 1995–2005, of interventions in Canadian child welfare. Four evaluations were based on randomized controlled trials, and six on non-randomized, quasi-experimental designs. After a critical review of each study, we formulated the implications of the review, for the design and evaluation of child welfare interventions in Canada, in terms of three main needs: for more high-quality impact evaluations; for the evaluation of the effectiveness of a wider range of interventions; and for the implementation and evaluation, in the Canadian context, of interventions of mainly U.S. origin that incorporate the principles of the increasingly influential perspective of evidence-based practice.

In 2004, an electronic survey of researchers and other stakeholders in child welfare in Canada was conducted to identify the most pressing areas for research (Vandermeulen, Wekerle, & Ylagan, 2005). The non-random sample of more than 200 respondents rated the evaluation of service effectiveness as the most urgent priority, in these terms: “How are current services working for clients, both in the short-term and across the lifespan?” (emphasis added; Vandermeulen et al., 2005, p. 3). The respondents also identified other research topics as important: evidence-based practice, prevention, permanency planning, the overlap of mental health and child welfare, and poverty in client families.

We decided to focus the present review on evaluations of program impact because stakeholders had identified service effectiveness as their top priority. After a brief review of child welfare in Canada, including recent trends in child maltreatment and out-of-home placement that make a focus on service effectiveness even more pertinent and timely, we provide a critical and detailed summary of the findings from four randomized and six quasi-experimental impact evaluations. We conclude with suggestions for increasing the quantity and improving the quality and relevance of such evaluations.

CHILD WELFARE IN CANADA

Organization of Services

In Canada, child welfare is a mandatory service, governed by legislation specific to each of the 10 provinces and 3 northern territories (Trocmé et al., 2005). In most jurisdictions, child welfare services are delivered by government employees, whereas in others (especially Ontario), independent agencies (e.g., nonprofit Children’s Aid Societies) provide services with government funding. In 2003, an estimated 76,000 children and adolescents in Canada were in out-of-home care in Canada (Faris-Manning & Zandstra, 2003), and many more were served by child welfare agencies while remaining in their families. Precise national data on the number of children served in out-of-home care or in their families are not available, because each jurisdiction
defines the terms — even those as basic as “child in care” or “child in need of protection” — in its own way. In the absence of uniform definitions, national data cannot be collected routinely, and national studies in child welfare in Canada are thus very rare. The only exceptions are specially funded national studies such as the recently published Canadian Incidence Study of Reported Child Abuse and Neglect (CIS-2003; Trocmé et al., 2005).

Recent Incidence and Trends in Maltreatment

CIS-2003 is a landmark investigation that provides a great deal of information on the “demand” (maltreatment) side of child welfare as well as some on the “supply” (service) side. The core CIS-2003 sample included all Canadian jurisdictions except Quebec (the recent implementation of a common information system in that province precluded its inclusion in the core sample). Trocmé et al. (2005) estimated that there were 103,297 substantiated child maltreatment investigations in Canada in 2003, for an incidence rate of 21.71 per 1,000 children aged 0–15 in the general population. (Following CIS-2003, the figures we cite here do not include Quebec.) Neglect was the most common form of maltreatment, accounting for 30% of all substantiated maltreatment investigations (a rate of 6.38 substantiated cases per 1,000), followed by exposure of children to domestic violence (28%, or 6.17 per 1,000), physical abuse (24%, 5.31 per 1,000), emotional maltreatment (15%, 3.23 per 1,000), and sexual abuse (3%, 0.62 per 1,000).

The rate of substantiated maltreatment increased in Canada by 125% between 1998 and 2003. This dramatic rise was due to changes in the way child welfare workers classify investigated cases, with relatively more classified as “substantiated” and fewer as “suspected” than previously. The increase also reflected more systematic investigation of siblings in families in which maltreatment of a target child was already being investigated as well as greater attention to emotional maltreatment (the incidence of which increased 276%) and exposure to domestic violence (which rose 259%). Regarding the other forms of substantiated maltreatment, physical abuse increased by 107% between 1998 and 2003, neglect increased by 78%, and sexual abuse decreased by 30%. The increases in the incidence rates for emotional maltreatment, exposure to domestic violence, physical abuse, and neglect were all statistically significant, whereas the decline in the incidence rate for sexual abuse was not (Trocmé et al., 2005).
Recent Trends in Out-of-Home Placement

Of the 103,297 substantiated child maltreatment cases in 2003, 8% (8,260) resulted during the initial investigation in placement of the child (or children) outside the home, in a formal child welfare setting: 5% in family foster care and 1% each in kinship foster care, group homes, and residential/secure treatment facilities. During the initial investigation, another 5% of the children were placed into informal kinship care, and out-of-home care was being considered as an option for a further 4%. Moreover, after the initial investigation period was over, further admissions to out-of-home care were likely but documenting them fell outside the temporal scope and mandate of CIS-2003 (Trocmé et al., 2005).

These CIS-2003 data on the growing number of substantiated maltreatment cases imply a concomitant increase in the demand for out-of-home placement. Such an increase is consistent with data indicating that the number of young people placed in out-of-home care in Canada has indeed risen recently. Trocmé and Chamberland (2003) reported that between 1996 and 2002, the number of children in out-of-home care increased by 60% in Ontario, 59% in Alberta, and 38% in British Columbia, and, between 1995 and 2001, by 71% among First Nations children.

METHOD

Inclusion Criteria

We define a program impact evaluation as an assessment of “whether a program produces desired effects over and above what would have occurred without the intervention or, in some cases, with an alternative intervention” (Rossi, Lipsey, & Freeman, 2004, p. 235). To be included in the present review, an evaluation had to have been published (or in press) in a peer-reviewed source within the last decade (1995–2005), because we wanted to cover relatively recent and higher-quality studies. The evaluation also had to have been an assessment of the impact of a program located in Canada and possessing an obvious relationship to the field of child welfare. Finally, the evaluation had to have been either a randomized controlled trial, with intervention and control groups formed through random assignment, or a quasi-experimental study, with intervention and comparison groups established non-randomly. We limited the evaluations to these two designs because they furnish stronger evidence than others of
program effects (Rossi et al., 2004). Of the two, however, randomized designs, if properly executed, provide a considerably stronger logical warrant than quasi-experimental designs for attributing desirable outcomes (or unintended negative effects) to the program itself rather than to extraneous influences.

Search Strategy

In our search for relevant impact evaluations, we tried to be as thorough as possible. First, we sent out an e-mail message, with one follow-up, to the 169 individuals listed in the database of Canadian child welfare researchers compiled by the Centre of Excellence for Child Welfare. We advised the researchers that we intended a broad definition of the term “evaluation” and invited them to send us paper or electronic copies of or references to evaluations of which they were the authors, whether their studies were in published, in-press, or unpublished form. Second, for the 1995–2005 period, we conducted an electronic search of PsychINFO, the Cochrane collaboration, Medline, and ProQuest journals, together with an electronic and manual search of 29 Canadian, American, and British journals in child welfare, social work, psychology, and psychiatry and a manual search of many book chapters.

We limited the review to the four randomized and six non-randomized published and peer-reviewed evaluations that we found among the more than 100 papers or chapters examined. In our critical summary of each evaluation, we have tried to provide enough detailed information so that the reader may understand the main features of the study and, short of reading the original paper, may assess the accuracy and fairness of our comments on it. Also, to avoid overburdening the text, we have chosen, in almost all cases, to refer to the concepts involved (e.g., self-esteem or anxiety) rather than to the specific instruments used to measure them.

RESULTS

Randomized Program Impact Evaluations

Prevention of Recurrence of Child Physical Abuse and Neglect

MacMillan et al. (2005), of McMaster University, published, in The Lancet, what in our opinion is the most rigorous impact evaluation of a Canadian child welfare program in the last decade. They in-
vestigated the effectiveness of a nurse home-visitation program in preventing the recurrence of physical abuse or neglect in families in which either form of child maltreatment had already occurred. All English-speaking families that had been referred to the only two child protection agencies in Hamilton, Ontario, during a 19-month period in 1995–1996 were eligible for participation in the study if an index child, younger than 13 years of age, had been exposed to physical abuse or neglect in the family within the previous three months and had also continued to live with the family. One hundred and sixty-three families were randomly assigned to either an intervention group (n = 89) or a control group (n = 74). Both groups received standard services arranged by the child protection agencies, consisting of follow-up services by agency case workers, such as assessment of recidivism risk, parenting education, and referrals to community-based parent education and other services. In addition, each family in the intervention group received a 1.5-hour home visit by a public-health nurse every week for the first six months, every two weeks during months 7–12, and once a month during months 13–24. Before intervening, the nurse-visitors received a week of manual-based training, and during the study they participated in group supervision meetings every two weeks. The nurses tailored their home visits to the needs of each family, providing intensive family support, education about child development, and linkages with relevant health and social services.

The main outcome consisted of physical abuse or neglect of any child in the family, including those born after randomization, based on protection-agency reports, during the three-year intervention period. Data from agency records were extracted and summarized by trained research assistants who were masked about the group to which the families had been randomized. Two experts in child maltreatment, also masked to group assignment, assessed the summaries of all visits or incidents to determine whether physical abuse or neglect had occurred.

Data analysis was based on “intention to treat,” a term that, although variably defined (Hollis & Campbell, 1999), is generally interpreted to mean the inclusion of all participants in a randomized intervention trial in the groups to which they were originally randomly assigned, “regardless of whether they actually satisfied the entry criteria, the treatment actually received, and subsequent withdrawal or deviation from the protocol” (Hollis & Campbell, 1999, p. 670). An intention-to-treat analysis is seen as essential in the evaluation of the effective-
ness of a “real-world” intervention, because clinical effectiveness may be overestimated if such an analysis is not carried out. Thus, in the study by MacMillan et al. (2005), two intervention-group families who, in fact, received no intervention were included in the analysis.

Cox regression and growth curve analysis were used in analyzing the data. At the three-year follow-up, in which 85% of the families participated, there were no significant differences between the intervention and control groups on the primary outcome, namely, recurrence of child physical abuse (33% in the intervention group vs. 43% in the control group) or neglect (47% versus 51%). Also, there were no significant differences between the groups on the secondary outcomes of parenting capacity, child behaviour, and quality of the home environment. Moreover, secondary outcome data from local hospitals unexpectedly showed that the intervention families had a higher reoccurrence of either physical abuse or neglect than did the control families (24% versus 11%).

MacMillan et al. (2005) felt they could rule out several potential threats to the validity of their findings: faulty randomization, low or differential follow-up rates in the intervention and control groups, differential detection of child maltreatment (including surveillance bias) in the two groups, and faulty implementation of the intervention program or control services. They concluded that their intensive two-year program of home visits by nurses was no more effective in preventing recurrence than standard follow-up services by caseworkers from the child protection agencies. They also concluded that the risk of recurrence is high when children remain in homes where abuse has already taken place, such that much greater efforts need to be put into prevention before a familial pattern of abuse or neglect has been established.

In the same issue of *The Lancet*, Barlow and Stewart-Brown (2005) had some interesting and insightful comments on the study by MacMillan et al. (2005) that have wider implications for impact evaluation in child welfare. They suggested that the study was one of the first rigorous evaluations of an intervention to stop child abuse in families already in the child welfare system and commended its use of an ecological theoretical model. They were also of the opinion that the higher rate of recurrence of abuse in the intervention group, based on data from hospital records, pointed to the existence of a surveillance bias. Further, Barlow and Stewart-Brown (2005) speculated that the success of intervention programs such as home visiting may depend on
the establishment of a trusting relationship between service providers and recipients and recommended that the measurement of this relationship should become a priority in such studies. They wondered whether trusting relationships had been at the core of the intervention of MacMillan et al. (2005) and also whether the program had had other key elements such as adequate intensity. Finally, Barlow and Stewart-Brown (2005) made an important observation about the inherent complexity of interventions such as that of MacMillan and her colleagues, an observation pertinent to the design and evaluation of virtually any program in child welfare, including those covered in the present review.

Trials that involve participants as active agents in initiating psychological and behavioural changes are fundamentally different from trials of treatments that involve the participants as passive recipients (as is the case for many medical interventions). Such trials are delivered by people with differing levels of skills, to families with different histories, preoccupations, circumstances, different levels of readiness for change, and different levels of commitment to the process, all of which determine whether an intervention of this nature proves to be effective. The implication of these differences is that the impact of such interventions is likely to be small, and there is, as such, a need for much larger (and more expensive) studies. Indeed, the survival curve in MacMillan and colleagues’ study shows a 10% reduction in physical abuse, which would seem to be an important improvement. However, the sample size was based on a 25% reduction, which might have been overly optimistic. (Barlow & Stewart-Brown, 2005, p. 1751)

Prevention of Child Sexual Abuse

Tutty (1997), of the University of Calgary, evaluated a sexual abuse prevention program known as Who Do You Tell, for children of elementary-school age. Tutty attempted to overcome some of the methodological weaknesses of previous research by using a relatively large sample, random assignment to conditions, and a standardized outcome measure, as well as by attending to age-related developmental differences. Two hundred and thirty-one children in grades 1 to 6 in two Catholic elementary schools in Calgary were randomly assigned to the prevention program (n = 117) or to a wait-list control condition.
The program was presented by two trainers to groups of 15–20 children, in two sessions of 45–60 minutes that took place on successive days. Prevention concepts were taught by means of age-appropriate pictures and videos, discussion, role-plays, and practice using the concepts, and the materials were matched to the children’s developmental levels. The program emphasized the communication of information and giving permission to say no to unwanted touch. The outcome measure was a 33-item measure known as the Children’s Knowledge of Abuse Questionnaire-Revised (CKAQ-R; Tutty, 1995), consisting of a 24-item Inappropriate Touch subscale (on which previous psychometric data existed) and a new, 9-item Appropriate Touch subscale (on which no such data were available). The CKAQ-R was administered verbally to the children by the author before and after the intervention, individually to children in grades 1 and 2 and in small groups of 5–15 to the older children.

On the Inappropriate Touch subscale, an analysis of covariance (ANCOVA) on the posttest scores, controlling for the pretest score, showed that the children exposed to Who Do You Tell scored significantly higher than the wait-list control children, as did older children and those with a higher score on the pretest. There was no gender difference nor were any of the interactions significant. An ANCOVA on the Appropriate Touch subscale posttest scores similarly revealed significant effects for treatment condition, age, and pretest score, and no gender or interaction effects. Of the 46% of parents responding to a questionnaire that sought demographic information on the families and reactions to the program, most were favourable toward the program, and few noted negative reactions on the part of their children.

Overall, Tutty (1997) interpreted her main finding — a statistically significant but relatively small increase in knowledge of prevention concepts — as consistent with most well-controlled research on sexual-abuse prevention programs directed at children which, as in the present case, are typically brief, of about 1–2 hours duration. Also, while acknowledging that children’s gains in knowledge constitute only a limited outcome criterion for evaluating sexual-abuse prevention programs, Tutty pointed out that behavioural-change criteria, in which children’s responses to strangers are typically assessed in the context of simulated potentially abusive situations, present very difficult ethical issues and do not cover the majority of real-world cases, in which the perpetrators of sexual abuse are known to the child-victim.
Prevention of Dating Abuse and Violence

Wolfe et al. (2003), of the University of Western Ontario and the Centre for Addiction and Mental Health and the University of Toronto, evaluated a community-based program for adolescents aged 14 to 16 years who were seen as at risk of developing abusive relationships with intimate partners because of their own histories of maltreatment. The program aimed at the prevention of abusive behaviour and the development of healthy, nonviolent relationships. Most of the maltreated youths (over 90%) were clients of one of seven child protection agencies in southwestern Ontario or Toronto, and the others were recruited from a special needs school. Youths were excluded from the study if they were receiving or were in more urgent need of other mental health services, had convictions for crimes against persons, or were developmentally delayed.

The 191 initial participants included 92 boys and 99 girls from mainly lower-income families. Two-thirds were randomly assigned to an intervention condition (Youth Relationships Project [YRP]) and the remaining one-third to a control condition (i.e., standard protection-agency services, consisting of bimonthly visits from a social worker and basic shelter and care). Subsequently, 25 (21%) of the intervention group dropped out (intervention dropouts were defined as those who had attended no more than five sessions), compared with 8 (11%) of the control group (dropouts from which were defined as those who had failed to complete an assessment beyond the initial intake). Analysis of the data was thus not based on the intent-to-treat principle, such that the final research sample was composed of 158 participants (96 in the intervention group and 62 in the control group). Completers and dropouts, both overall and within the intervention group, did not differ on maltreatment history or on the outcome measures.

The intervention (YRP) is a program that involves 18 two-hour sessions and employs a health-promotion approach to the prevention of violence in dating relationships. YRP focuses on positive alternatives to aggression in resolving interpersonal problems, and the curriculum has three components: education about abuse and power in intimate relationships, skill development, and social action activities (e.g., visits to social agencies). A total of 15 coeducational intervention groups were conducted, each involving 6–10 participants and led by adult male and female facilitators who had experience with youths or with domestic violence. Each of the 18 sessions was described in a YRP program manual. A fidelity check indicated that the program
was implemented as intended, and several intervention process variables were assessed (e.g., the number of treatment sessions attended; the level of participants’ listening skills, disruptive behaviour, and involvement; and the overall quality and cohesion of the group process). The outcome measures assessed abuse and victimization with dating partners, emotional distress, and healthy relationship skills. On average, the participants were followed up for 16 months and assessed 4.7 times.

Growth curve analyses of change on each of the outcome variables indicated that, as hypothesized, participants in the YRP intervention, compared with those in the control condition, were less physically abusive toward their dating partners. They also reported receiving less physical and emotional abuse and experiencing fewer threats and also (although this was not a direct goal of the intervention) exhibited a greater decline in symptoms of emotional distress. There was no differential impact on healthy relationship skills. Wolfe et al. (2003) concluded that the positive results of the intervention were due to the intervention but could not rule out the possibility that they may also have reflected non-experimental factors (e.g., greater awareness of the target behaviours and increased self-monitoring). Interestingly, of five intervention-process variables examined in the analyses (number of treatment sessions attended, and co-facilitator ratings of each participants’ listening skills, disruptive behaviour, involvement, and group cohesion), only listening skills were related to a decline in abusive behaviour. Unexpectedly, involvement in the intervention group was associated with a trend toward an increase in abusive behaviour. Thus, the group format of the YRP program may have produced iatrogenic effects, at least in some young people.

Enhancement of Parenting Among Maltreating Mothers

Hughes and Gottlieb (2004), of Dalhousie and McGill Universities, conducted a randomized controlled trial of the effect of the Webster-Stratton (1989) parenting program on positive parenting strategies and child autonomy within maltreating families. According to Hughes and Gottlieb, the Webster-Stratton program is the only group-based parenting program with demonstrated effectiveness among high-need families facing problems known to be associated with child maltreatment, including children with conduct disorder, poverty, and living on social assistance. A standardized, video-based, modelling program based on social learning theory, the program is geared to the developmental needs of families with young children, aged 2–8 years.
In what was apparently the first evaluation of the Webster-Stratton program in maltreating families, Hughes and Gottlieb (2004) predicted that, compared with mothers and children in a wait-list control condition, maltreating mothers exposed to the intervention would show higher levels of three positive parenting behaviours: autonomy-support (i.e., enhancing the child’s sense of value and personal control), structure (enhancing the child’s mastery by setting limits and boundaries), and involvement (showing praise, nurturance, and appreciation of the child). Also, the children exposed to the Webster-Stratton program were expected to exhibit greater autonomy (i.e., an amalgam of curiosity, interest, optimal challenge, persistence, independent decision-making, independent mastery, willing cooperation, and persuasive resistance).

Following a baseline assessment, 14 maltreating families (i.e., mothers and their children) on the active caseloads of three child protection agencies in Eastern Canada were randomly assigned to the Webster-Stratton program, and 14 were assigned to a wait-list control group. The intervention was an additional service received, not a substitute for other services. More than half of the mothers and nearly half of the children in each condition were receiving other mental health services. The mothers were active clients of a child protection agency, were considered at high risk of committing abuse and thus in need of parent training, spoke English, had completed at least grade four in school, and had freely consented to participate. The target children were, in almost all cases, between 3 and 8 years of age, did not have a severe mental disability, and lived with their mothers. Of the 28 participating mothers, two dropped out, one from each condition. Twelve of the 13 remaining Webster-Stratton program families received six or more of the eight planned parenting sessions. The measures of parenting skills (autonomy-support, structure, and involvement) and child autonomy consisted of ratings made by research assistants who, masked to the experimental condition, rated video-taped mother-child interactions during pre- and post-intervention assessments.

Analysis of the data was not based on intent-to-treat. An ANCOVA showed that the mothers exposed to the Webster-Stratton parenting program were significantly ($p = .03$) more involved with their children than those in the wait-list condition on the parenting skill of involvement. There was a trend toward a program effect on autonomy-support ($p = .07$), but no difference between groups on structure. The program had no impact on the children’s autonomy. Overall, the relatively weak intervention effects suggested that had the program
focused on both mothers and children and their real-world interaction, it might have had a larger impact.

It is interesting that, despite its weak effects, the intervention elicited a high degree of participant satisfaction. The mothers reported gains in their own self-esteem, knowledge of child development and parenting strategies, improvements in familial harmony, and increases in the child’s attention, listening skills, and happiness. In our opinion, however, we think that these subjective estimates of outcomes should be interpreted with a great deal of caution, given the modest objective effects (as rated by the masked research assistants) and the possibility that these subjective judgements may have reflected strong demand characteristics created by the intervention itself.

Non-Randomized, Quasi-Experimental Program Impact Evaluations

Prevention of Child Neglect in At-Risk Families

Ethier, Couture, Larcharité, and Gagnier (2000), of the University of Québec in Trois-Rivières, evaluated the short-term effects of a multidimensional intervention program designed to prevent child neglect, defined as chronic failure on the part of parents to meet the basic needs of a child in the physical, social, affective, or cognitive domains. In Quebec, as in other Canadian jurisdictions, neglect is a relatively frequent form of child maltreatment. Risk factors common in child-neglecting families include low family income and education, single parenthood, marital violence, mother aged less than 21 at the birth of the first child, physical abuse or neglect in the parent’s own childhood, perinatal complications such as premature birth or very low birth weight, substance abuse, mental health problems, and high levels of depression and stress (Ethier et al., 2000). Given this multiplicity of risk factors, the research team adopted an ecological perspective, with a simultaneous focus on several levels: parent-child relationships, the family, and family-environment interactions. They adapted a program known as the Personal, Family, and Community Help Program (PFCHP), originally designed for maltreating families, to the prevention of neglect in at-risk families.

The PFCHP lasted 18 months and involved the formulation of individual objectives that were coordinated by a clinical team that met every week. The ecologically-oriented program approach comprised four main elements. First, adults working in community services were recruited, trained, and supervised to act as volunteer visiting
families who would phone and visit up to three at-risk families on a twice-a-week basis. Their task was to act as parental role models and to provide the target families with support in relation to accessing community services, such as medical appointments and school, and daily problem-solving, such as meal planning or choosing appropriate clothing for their children. Second, weekly group meetings, held over a 44-week period and led by an experienced therapist, were attended by the at-risk families, focusing on relational and parental competency issues, to increase the parents’ awareness of their own psychological needs, sensitivity toward their children, and problem-solving skills, and to decrease their parental stress and anxiety. Third, educational activities for the children of the at-risk parents, directed by two special-education graduate students, were held at the same time as the group meetings, to stimulate the children’s language, cognitive, and social skills. Fourth, “on-demand” individual counselling was provided to each family by the social worker assigned to that family. The PFCHP was compared with the conventional social services offered to at-risk families in typical Local Community Services Centres (LCSC). The LCSC approach focused primarily on the enhancement of parental skills and the parent-child relationship, coupled with referrals to appropriate community services. Meetings with the family’s social worker occurred every two weeks, on average, except when family crises required more frequent meetings. The LCSC program lasted from 12 to 20 months, with a mean of 16.3 months.

The evaluation design was quasi-experimental in nature. A total of 29 families took part, 15 PFCHP and 14 LCSC families. The PFCHP families, recruited through a local LCSC in the Trois-Rivières area of Quebec, had at least four socio-demographic, relational, or medical risk factors, had never been clients of youth protection services, and had at least one child under 6 years of age. The LCSC families were recruited when they sought services for the first time at the same LCSC. The PFCHP families were recruited first, followed by the LCSC families. Fifteen of the 16 PFCHP families completed a pretest and postevaluation, and 12 of 14 LCSC families did so. Data analyses were carried out on the mothers’ data only, as only 6 men completed both the pretest and the posttest assessments. (In this evaluation, as in the other five employing a quasi-experimental design, intent-to-treat was not used. However, the logic of this approach—now seen as essential in randomized evaluations of the “real-world” effectiveness of interventions—strikes us as also applicable to non-randomized evaluations, in which a failure to include intervention or compari-
son group drop-outs would simply add to the interpretive ambiguity already inherent in quasi-experimental designs).

The mean age of the mothers was 20 years, and there were from 1 to 4 children in the mainly low-income families. The two groups of families appeared to be equivalent on socio-demographic grounds, and each group was exposed, on average, to the same number of risk factors (7.6). Standardized psychometric measures were used to assess the size of the mother’s social support network, her satisfaction with the support received, and the parental level of stress, depressive affect, and child abuse potential. Also, an interview was conducted to investigate the evolution of the family during the intervention process, information supplemented by the field notes of social workers and visiting families and by case study reports. These qualitative data were organized into broad categories and later rated, by judges masked to the family’s membership in the intervention versus comparison group, as evidence of progress, stability, or regression in each category (e.g., parent-child relationships or social support).

Repeated-measures ANOVAs, used to analyze the data, produced little or no evidence that the PFCHP social-ecological intervention was any more successful than the conventional LCSC social-service intervention for preventing child neglect. That is, on the outcomes of social support, parenting stress and depression, and potential for child abuse, very few of the relevant statistical tests (i.e., the groups-by-times interaction F-tests) were statistically significant. Moreover, on the two subscales of the child abuse potential inventory where significant interactions were found, application of a Bonferroni correction for multiple analyses (Stevens, 1986) would almost certainly have eliminated these as well.

Overall, the mothers in both groups showed, on average, equal increases in their social-network satisfaction, as well as equal decreases in their level of parental stress, depressive affect, and child abuse potential (which, however, remained relatively high). The qualitative data suggested that the PFCHP intervention was related to improvements in the mother’s disciplining of her children, involvement in their activities, relationship with her partner, social support, and return to school or work, whereas the LCSC services were associated with gains in the mothers’ self-confidence and awareness of appropriate parental attitudes. Limitations of this study included the small number of families in each group and consequent limited power, the possibility that some of the gains observed in the two groups were
testing or demand-characteristic effects rather than program effects, the focus on short-term results, and the omission of a measure of actual (as opposed to potential) child neglect.

Group Treatment of Men Who Have Experienced Childhood Sexual Abuse

Tourigny, Guillot, and Morissette (2005), of the Universities of Sherbrooke and Montreal, evaluated the effects of a group intervention for men who had been victims of childhood sexual abuse. Group interventions seem especially popular in serving male child sexual abuse victims because of their putative power to overcome isolation and stigma and reduce the negative effects of the abuse. Few evaluations of the effectiveness of such interventions have been conducted, however.

In the quasi-experimental evaluation by Tourigny, Guillot, et al. (2005), 26 participants who met several inclusion and exclusion criteria and consented to participate formed the intervention group. They were recruited among men referred (or self-referred) to a community agency offering services to men who been sexually abused during childhood. Of these, 1 did not return the pretest materials on time and 2 dropped out during the study, leaving 23 in the intervention group. The non-randomly constituted comparison group was composed of 17 men who met the same inclusion criteria as well as the additional one of not receiving psychotherapy during the period of the study. They had been recruited via newspaper advertisements, notices on university or community-agency bulletin boards, or from the waiting list of the community agency from which the intervention group had been recruited. Of the 17 who initially agreed to participate, 2 were subsequently excluded because they entered therapy and 2 dropped out during the project, leaving 13 in the comparison group. Overall, the level of childhood sexual abuse had been quite severe in both groups.

The intervention evaluated by Tourigny, Guillot, et al. (2005) was psycho-educational in nature, composed of 10 weekly 3-hour sessions that were limited to a maximum of 8 group members and led by a team of male and female facilitators. The goals were to break the sense of isolation experienced by participants, increase their sense of personal worth, and attenuate the effects of the past abuse on their adult functioning. Pretest and posttest measures included a questionnaire used to collect socio-demographic data and information on the
nature of the sexual abuse and on the perpetrator, and standardized measures of post-traumatic stress, symptoms of psychological distress, self-esteem, and loneliness.

The pretest data showed that the two groups were rather dissimilar, with those in the intervention group significantly less likely to be employed, more likely to have been abused for 12 months or longer, displaying higher mean levels of post-traumatic stress, psychological distress, and loneliness, and a lower mean level of self-esteem. The relevant statistical tests, the groups-by-time interaction $F$-tests, were significant, at the $p < .05$ level, for only 2 of the 10 outcomes evaluated, namely, the traumatic effects of sexual abuse and self-esteem. Although the authors interpreted these results as supportive of the effectiveness of the intervention, a more conservative analysis, with application of the Bonferroni correction (Stevens, 1986), would have led to a different conclusion. That is, had each interaction-$F$ been tested at the $.05/10 = .005$ level, neither effect (involving, respectively, the traumatic impact of sexual abuse and self-esteem) would have been statistically significant. Thus, in addition to the limitations of the study noted by the authors themselves — small sample size, non-equivalence of the groups, possible regression to the mean in the intervention group, possible influence of extrinsic events, and lack of a post-intervention follow-up — the most parsimonious interpretation of the results would seem to be that the intervention had no significant effects.

**Group Treatment of Adolescent Girls Who Have Experienced Sexual Abuse**

Tourigny, Hébert, Daigneault, and Simoneau (2005), of the Universities of Sherbrooke and Québec at Montreal, evaluated the effects of a treatment group for adolescent girls who had experienced sexual abuse. In an attempt to remedy some of the problems noted in earlier research, the authors included a comparison group, a wide range of measures, including some specific to sexual abuse, a larger than usual sample, and an assessment of participants’ level of involvement in treatment sessions.

The authors’ quasi-experimental pretest posttest design compared 27 teenage girls (mean age at pretest = 14.6 years) in a closed-therapy intervention group who received weekly 2-hour sessions for 20 weeks from a Centre d’Intervention en Abus Sexuel pour la Famille (CIASF) with a comparison group of 15 teenage girls (mean age = 14.8 years).
who had requested but did not receive services from CIASF. Members of the comparison group received no services because they had finally decided not to take part in the group therapy, had withdrawn from the group before the third session, or because, in the judgement of the clinical staff, group therapy was unsuitable for them (e.g., because of their timidity or difficulty in expressing emotion in a group context). The sexual abuse had been severe in most cases. The intervention and comparison groups were similar on socio-demographic, perpetrator-related, and abuse-related variables, except that a higher proportion of girls in the comparison group had experienced an episode of abuse with penetration (85%) than those in the intervention group (50%). The study began with 30 girls in 4 different intervention groups. Three dropped out before the third session and were included in the comparison group. The 27 who remained attended, on average, 90% of the group-therapy sessions ($M = 17.3$). The two groups were similar at the pretest, except that the comparison group had greater difficulties with eating-related disorders, interpersonal trust attributions, and coping via social-support seeking and planful problem-solving.

The CIASF is a community agency in Quebec that offers closed group therapy for 13- to 17-year-olds referred by community agencies. The therapy program evaluated is “closed” in the sense that new participants are not accepted once a group has started. The intervention is psycho-educationally oriented, follows a treatment manual, is led by two facilitators (female and male) who have at least an undergraduate degree in social work, psychology, or sexology, and is given to groups of 6–8 in 20 weekly 2-hour sessions. The objectives are to address the negative psychological consequences of sexual abuse, reduce abuse-related social isolation, shame, or guilt feelings, and help participants use their own resources and skills in managing the consequences of abuse.

The standardized outcomes assessed consisted of measures of participants’ self-reported trauma symptoms, behavioural problems, coping strategies, empowerment, sexual abuse-related self-perceptions and attributions, self-injurious behaviours, and perceptions of the quality of their relationships with their mothers and fathers. Two measures used only at the pretest were a self-report socio-demographic questionnaire and a practitioner-administered rating scale of the sexual abuse (severity, frequency, perpetrator identity, etc.).

In the groups-by-times interaction analyses for the 8 outcomes involving post-traumatic stress symptoms and child attitudes toward
the parents, 6 of the 8 interaction-\(F\)s were significant at the \(p < .05\) level or better (excluding the interaction-\(F\) for the total score on one instrument that was largely redundant with the 6 interaction-\(F\)s calculated for its subscales). However, had a Bonferroni correction been used, such that each interaction-\(F\) would have been tested at the \(p < .00625\) level, only 3 (rather than 6) of the 8 would have been statistically significant (i.e., those for the symptom dimensions of anxiety, depression, and post-traumatic stress).

Similarly, in the analyses for coping, empowerment, and attributions, of the 8 non-redundant interaction-\(F\)s tested, a Bonferroni correction would have found that only 3 (rather than 5) were significant, at the \(p < .00625\) level, namely, those for coping via social support, interpersonal trust, and personal attributions of negative events. Also, of the 8 non-redundant self-reported behavioural-problem scales, the Bonferroni procedure would have found only 3 (rather than 4) program effects significant at the \(p < .00625\) level, namely, improvements in anxiety, social withdrawal, and attention. Finally, in the analyses for the 3 outcomes of eating disorders, self-harming behaviours, and delinquent behaviours, none of the interaction-\(F\)s were significant even when each was tested at the liberal \(p < .05\) level, much less at the Bonferroni-corrected \(p < .05/3 = .017\) level.

Overall, Tourigny, Hébert, et al. (2005) were able to establish as plausible that the group therapy produced some effects, although the number of such effects was smaller than claimed. The study had several limitations shared by many of the other quasi-experimental evaluations reviewed: small sample size, a single source (i.e., the participants themselves) of data on outcomes, and a non-conservative data analytic strategy. The authors did, however, conduct a careful implementation analysis of the proportion of planned group therapy sessions in which the adolescent girls had actually participated (on average, 90%).

Group Treatment of Adult Women Survivors of Childhood Sexual Abuse

Westbury and Tutty (1999), of the University of Calgary, used a quasi-experimental pretest/posttest comparison group design to evaluate the effectiveness of a new group treatment for adult women survivors of childhood sexual abuse, compared with individual therapy alone. The group therapy, a feminist model developed at the Calgary Family Service Bureau, was based on Integrative Body Psychotherapy, a
holistic approach that considers cognitive, emotional, physical, and spiritual aspects of the survivor’s life to be important. The closed therapy groups, composed of 6–8 members and led by 2 therapists, meet weekly for 2.5 hours for 10–12 weeks. Each participant must previously have undergone 6 months of individual therapy focused on issues of childhood sexual abuse and may also engage in individual therapy during the group therapy.

In the present study, the intervention group was composed of 22 women survivors of severe childhood sexual abuse who were receiving or wished to receive treatment at the Calgary Family Service Bureau. All were concurrently receiving individual therapy focused on childhood sexual abuse concerns and had already received at least 6 months of the same kind of individual therapy. The 22 participants were seen in 5 treatment groups. The comparison group consisted of 10 women on the agency waiting list for treatment. The women survivors selected themselves into the intervention or comparison group, based on convenience or a preference for a particular facilitator team. Standardized instruments were used to assess the outcomes of depression, self-esteem, and trauma symptoms, after the first group session (pretest) and prior to the last group session (posttest). There were no significant differences between the groups on any of the pretest measures.

ANCOVA was used to compare the posttest scores of the members of the intervention and comparison groups, controlling for their pretest scores. A total of 7 ANCOVAs were conducted, one on each of the outcome measures (excluding the total score for the trauma measure, which was largely redundant with its 5 subscales). The groups were significantly different on only 2 of the 7 adjusted outcome means, those for depression \( (p = .009) \) and anxiety \( (p = .049) \). Had the conservative Bonferroni procedure been applied, however, with each ANCOVA tested at the \( .05/7 = .007 \) level, neither of the pairs of adjusted posttest means would have been significantly different. The most parsimonious interpretation of the findings would appear to be that the group therapy had no impact.

Intensive Family Support to Prevent Out-of-Home Placements

Dagenais et al. (2003), of the Universities of Quebec at Montreal and Sherbrooke, carried out an implementation evaluation, followed by a quasi-experimental impact evaluation, of a program of intensive support offered to families thought to be at risk of having one or more
children placed in out-of-home care. The program, known as PRIME (Project d’intervention massive à l’enfance), was operated by the Centre jeunesse de Montréal and based on the well-known Homebuilders program, which has inspired many such programs (often called intensive family preservation) in the US, Canada, and Europe. Such programs offer rapid, in-home, intensive support aimed at helping the family overcome an immediate crisis, improve its own and the target child’s psychological and social functioning, ensure the child’s security, and avoid an unnecessary placement of the child. Like other such programs, the services provided by PRIME were supposed to be rapid (typically offered within 24 hours of referral), intense (with a single worker assisting only 2 or 3 families), accessible (available on a 24-hour-a-day basis, 7 days a week), offered in the family environment (in the family home, school, or neighbourhood), collaborative (program staff work with staff from other agencies), brief (of 4–6 weeks’ duration), family-focused (involving all members of the family), and concrete (offering assistance with transportation, food, keeping appointments, etc.).

The PRIME evaluation lasted for more than 4 years and included an assessment of its implementation, effects on placement, and effects on the child and family. The implementation evaluation investigated whether PRIME had adequately implemented the key elements of the Homebuilders program. A total of 69 families, including 88 children who were at risk of out-of-home placement, participated in the implementation evaluation. The characteristics of the services provided by PRIME during 1995–1997 (the period during which the implementation and impact evaluation by Dagenais et al. [2003] took place) were compared with the services offered by PRIME during its first two years of operation, 1993–1995 (during which an earlier implementation evaluation had been conducted). Dagenais and his colleagues found that by the end of the four-year period (1993–1997), PRIME had been only partially implemented. Although services were offered in the family environment, collaborative, and brief (lasting only 4–6 weeks), they did not meet other Homebuilders criteria, namely, those related to involvement of the whole family, availability, concrete support, rapidity of initial response, and intervention intensity.

The evaluation of the impact of the program on placement and on the psychological and social status of the child and family employed a quasi-experimental design. The researchers had originally intended to compare the 88 PRIME children with a comparable number (80) who, though eligible for the service, could not be accommodated because of
a lack of capacity in PRIME. Numerous unforeseen administrative changes, however, limited the actual comparison group to 21 children, all of whom received regular services from the Centre jeunesse de Montréal. In supplementary analyses, Dagenais et al. (2003) matched the 21 comparison children with the 21 of the 88 PRIME children who most resembled them, on 9 variables: child age, sex, and ethnicity; placement episodes in the two years preceding the target request for placement; number of days placed before the request for services; type of placement; family structure and income; and the number of children in the family at risk of being placed.

The quasi-experimental impact evaluation showed that 26% of the 88 PRIME children had been placed outside the home within 3 months of the request for services, a placement rate that had grown to 44% at 6 months and 55% at 9 months. For the 21 matched PRIME children, the proportions were 24%, 43%, and 47%, respectively, no different from the proportions (43%, 52%, and 62%) observed for the 21 comparison children (who had received regular services). Regarding the number of days placed, the 21 PRIME children, on average, had been placed for fewer days during the first 3 months ($p = .003$), but this advantage had disappeared at 6 and 9 months. Finally, there was no significant difference in the amount of time that had elapsed between the request for services and the first placement. In sum, there was no evidence that PRIME had had any effect on placements.

The evaluation of the impact of the program on the children and families was based on 53 of the 88 PRIME children, namely, those whose families had received services from PRIME but who had not been placed. Within-group analyses compared the situation of the children and families at the time the request for services was made and 6 months later. Between-group analyses compared 40 children aged 10 and over in 38 families that had received services from PRIME with 18 children aged 10 and over in 18 families that had received regular services.

$T$-tests conducted within the PRIME group showed improvements ($p < .05$) over the first 6 months of services on most aspects of the child's behaviour, parental well-being, and a few aspects of family functioning. No improvements were observed, however, on placement risk factors, family social support network components (except for a significant decrease in the number of workers present within the network), or most aspects of family functioning, and the level of excessive control within the family unexpectedly increased significantly.
It seems likely that some of these \( t \)-tests would not have survived application of the Bonferroni procedure, a possibility that we could not evaluate because these results were summarized in verbal rather than in quantitative form.

The comparisons between the PRIME and regular-service users between the beginning of services and 6 months later showed no significant differences except for two (possibly non-replicable) exceptions: parents receiving regular services, compared with those receiving PRIME services, experienced a decrease in financial problems but an increase in mental health difficulties. Finally, there were no differences between the regular and PRIME groups in terms of the time elapsed between a request for services and a subsequent investigation, nor in terms of the average number of investigations per child.

In sum, Dagenais et al. (2003) found that PRIME had been implemented only partially and that its effects were no better than those produced by regular services. The authors interpreted these disappointing results as further evidence that, in line with earlier warnings from a previous implementation evaluation of the first two years of the project (1993–1995), PRIME had drifted away from the Homebuilders model towards a less intensive and more traditional approach. According to Dagenais et al., this drift was due, in part, to the fact that only the first PRIME team leader had received Homebuilders-based training. None of this leader’s successors had ever received Homebuilders training and were thus incapable of providing appropriate clinical supervision or project leadership. The researchers also pointed to other elements whose presence would have improved the program and its evaluation: a larger sample and greater statistical power; less heterogeneity in family difficulties and service objectives; greater clinical expertise on the part of staff; fewer organizational, leadership, or personnel changes; and an effective referral process that would have enabled the establishment of an adequate comparison group.

Dagenais et al. (2003) felt that the immediate family crisis that is the raison d’être of the Homebuilders approach was likely to have passed by the time the family had gone through all the steps (e.g., intake, investigation, emergency services, or assessment) preceding the receipt of services from PRIME. They also stated that a well-validated instrument capable of predicting which children in families in crisis were likely to be placed in out-of-home care was lacking. In this latter connection, it is worth underlining that the failure of research to address such gaps in our practice knowledge is costly in
service-effectiveness terms. For, more than a decade ago, in a widely used evaluation text, Wholey (1994) identified essentially the same program-design problem:

The evaluability assessment [of Kaye and Bell, 1993] concluded that current family preservation programs were not consistently targeted at families with children who were at imminent risk of foster care placement and that, as a result, the primary goal of policymakers could not be achieved as these programs were currently operated. (p. 30)

Mutual Aid Among Parents in Families Served in Child Welfare

Cameron and Birnie-Lefcovitch (2000), of Wilfrid Laurier and Memorial Universities, presented the outcomes for the Parent Mutual Aid Organizations (PMAO) in Child Welfare Demonstration Project. The PMAO model was based on the assumption that mutual aid organizations may be a promising and practical way of providing informal social support to parents who, as clients of child welfare, are often socially isolated and stigmatized. In the project, 3 parent mutual aid organizations (PMAOs) were created in 3 separate communities in Ontario, under the auspices of the respective local Children’s Aid Societies. Each PMAO was staffed by a program development worker, and parents, as voluntary members, were actively involved in administering the organization. The model emphasized the following membership benefits: a high level of direct contacts among members, 2–5 times per week, in scheduled activities or informal meetings; access to helpful activities, such as parent relief, personal development courses, emotional support, or concrete resources; a safe place and positive peer network; aid in forming friendships with other members; the opportunity to become a giver as well as receiver of assistance; responsibility for running the organization; and participation in a broad range of social involvements and successful social roles, such as group leader, volunteer, student, or employee. Each PMAO offered a wide range of experiences, beginning with recreational activities and then moving to a greater emphasis on members’ personal growth (e.g., cooking; fitness; parenting; academic and employment skills), management of their own organizational affairs, interpersonal contacts, and friendships.

The quasi-experimental evaluation included three sets of interviews, with 53, 81, and 97 PMAO members and 56, 60, and 58 comparison group members, respectively. The interviews took place on three occa-
sions, 12, 18, and 24 months after the PMAOs had been initiated. The comparison group was composed of randomly selected cases from the current child protection caseloads of the three host agencies. A high proportion of the PMAO and comparison group members interviewed at time 1 were re-interviewed at time 2, and a high proportion of those interviewed at time 2 were re-interviewed at time 3. Standardized measures of social support, self-esteem, group social environment, and problems and concerns, as well as open-ended qualitative questions, were administered in face-to-face interviews by trained interviewers. Focus groups were also conducted at each site with all available PMAO group members in the first and third years of the project and with program development staff, and additional qualitative data were also collected throughout the project.

The results suggested that both the PMAO and comparison group samples were coping with many day-to-day difficulties. However, the PMAO sample reported a higher level of problems and concerns (e.g., higher unemployment, less money for daily living expenses, less satisfaction with neighbourhood resources for their children, greater social isolation and loneliness, etc.) and lower social support, self-esteem, and ability to cope with stress. Across the 3 PMAO sites, about 50% of the parents referred to the project became involved, participating on average 1–3 times a week in formal activities and once a week in informal contacts with members outside of organized activities. Not surprisingly, the local context proved important, with the site with the most accessible transportation also developing the most active membership base.

On the key outcome variable of out-of-home placements, the researchers compared, at times 1, 2, and 3, those PMAO and comparison group members who had been present from the beginning of the study. Similarly, those who had joined the study at time 2 were compared at times 2 and 3, and those who had joined after time 2 were compared only at time 3. Cameron and Birnie-Lefcovitch (2000) claimed that chi-square analyses, conducted at each of these 6 time points, showed that the PMAO members had made significantly less use of child placements than did the comparison group on 5 out of 6 occasions. However, of the 6 chi-square tests in question, 3 were statistically significant only at the liberal level of $p < .10$, 2 were significant at the $p < .05$ level, and only 1 was significant at the $p < .01$ level. Thus, had the Bonferroni correction been applied, with each chi-square tested at the $\frac{.05}{6} = .0083$ level, only 1 (and perhaps none) of the 6 placement rate differences would have been statistically significant.
Participation in structured community activities was much higher among PMAO members than among the comparison group members. At the 12-, 18-, and 24-month time points, 84%, 89%, and 94%, respectively, of the PMAO members took part at least once a week in organized social or recreational activities, versus 10%, 23%, and 15% of the comparison group. These differences were highly statistically significant, even with the relatively small sample sizes involved, and would have easily survived an application of the Bonferroni procedure. The PMAO members were also considerably more likely to volunteer their time in the community at least once a week (31%, 61%, and 76%, vs. 3%, 12%, and 9%) or belong to a church or community group that they attended at least once a week (88%, 91%, and 80%, vs. 19%, 8%, and 15%), differences that were also highly statistically significant. Much of this considerably higher level of social integration was mediated by the PMAO organizations. On the other hand, at the three measurement occasions, the PMAO members were no more likely than the comparison group members to have gotten together with a close friend, one or more times a week (59%, 82%, and 71%, vs. 75%, 80%, and 68%).

Rather than employing the standard statistical test for the quasi-experimental comparison group design, namely, the groups-by-times interaction $F$-test obtained from a repeated-measures analysis of variance (ANOVA), Cameron and Birnie-Lefcovitch (2000) used multiple $t$-tests across occasions, within each group separately, to evaluate whether PMAO members ($n = 46$) were more likely than comparison sample members ($n = 47$) to have reported increases in social support (appraisal, belonging, tangible, and total support) across the three assessment occasions. They found no significant increases in the comparison group, versus 7 significant increases in the PMAO group (these increases were statistically significant, however, only at the very liberal $p < .10$ level). Once again, had a Bonferroni correction been applied within the PMAO group, with each evaluated at the $\frac{.05}{12} = .0042$ level, none of the 4 increases found to be significant at the .01 level might have survived (we could not tell, as the exact observed $p$-levels were not reported), and certainly none of the 3 increases in the PMAO group that were significant at the .05 level would have survived. Thus, despite the authors’ claims to the contrary, it is not clear that the PMAO project actually conferred any advantage over the comparison group in terms of social support. The absence of such an advantage would not be surprising, perhaps, given that PMAO and comparison group members had similar frequencies of contact with a close friend once a week or more often.
Similar comments can be made about the authors’ analyses of changes in self-esteem and perceived stress, which were analyzed with multiple $t$-tests rather than with interaction-$F$s from repeated-measures ANOVAs. Had separate Bonferroni corrections been used for each outcome measure within the PMAO group alone, neither of the increases in self-esteem that they found in the PMAO group would have been significant (at the $0.05/3 = 0.017$ level), and only the drop in the level of stress between time-1 and time-2 would have survived the procedure. Moreover, had a single Bonferroni procedure been applied to all 6 $t$-tests within the PMAO group for the self-esteem and stress outcomes combined (an arguably more defensible approach), then even the decline in stress might not have been found to be significant. These same comments can also be applied to the authors’ analysis of changes in parental attitudes, reported in less detail. The improvements in the PMAO group appeared very slight, and none may have survived the Bonferroni procedure. Finally, the authors’ suggestion that a PMAO model would save $420–$869 per case, or between $16,800 and $34,760 if 40 families were actively involved, appeared highly speculative. Overall, due to weaknesses in their statistical analyses and the inherent uncertainty of results derived from a quasi-experimental design, many of Cameron and Birnie-Lefcovitch’s (2000) assertions about the alleged benefits of their PMAO program are questionable, going beyond what their data will support.

DISCUSSION OF IMPLICATIONS

We return now to our initial issue of service effectiveness and ask, What are the main implications of our review for improving the design and evaluation of child welfare interventions in Canada? We frame our answer to this question in terms of three main needs: for more high-quality impact evaluations; for evaluations of the effectiveness of a wider range of interventions; and for the implementation and impact evaluation, in the Canadian context, of interventions — many of U.S. origin — that incorporate the principles of the new and rapidly growing perspective of evidence-based practice.

Need for More High-Quality Impact Evaluations

It is striking that despite our considerable efforts to identify as many impact evaluations as possible, we were able to find only 10 peer-reviewed studies (4 randomized, 6 non-randomized) that had been published during the 1995–2005 decade. Thus, it would seem that
the first order of business for child welfare researchers in Canada should be to increase the sheer number of impact evaluations published each year, beyond the current rate of about 1 per year, and to ensure that these are as high-quality as possible. In many of these evaluations, multiple sites will be needed, to ensure the enrollment of a sufficient number of participants and an adequate level of statistical power (i.e., .80 or higher). In our review, the problem of overly small samples and low statistical power was found in virtually all the quasi-experimental studies and even in one of the randomized studies. This issue of sample size and statistical power is even more crucial in evaluations of child welfare interventions, given that effect sizes are often likely to be only small or medium in size (Hollis & Campbell, 1999; MacLeod & Nelson, 2000).

Some of the impact evaluations reviewed had clear methodological strengths, including the use of randomized designs (Hughes & Gottlieb, 2004; MacMillan et al., 2005; Tutty, 1997; Wolfe et al., 2003); power and intent-to-treat analyses (MacMillan et al., 2005); relatively large samples (Tutty, 1997; Wolfe et al., 2003); sophisticated statistical analyses (MacMillan et al., 2005; Wolfe et al., 2003); an assessment of the adequacy of implementation as a prerequisite for the interpretation of program effects (Dagenais et al., 2003; MacMillan et al., 2005; Tourigny, Hébert, et al., 2005); attention to the pre-intervention comparability of groups (Ethier et al., 2000); and explicit consideration of the possible confounding role of non-experimental factors (Wolfe et al., 2003). On the other hand, a number of the evaluations had important methodological weaknesses. Besides the problem of inadequate sample size that has already been mentioned, few of the evaluations used effect-size measures to convey the strength of statistically significant results, and several of the quasi-experimental studies employed clearly non-comparable groups. Also, the failure in several studies to use standard statistical procedures, including, for example, repeated-measures groups-by-occasions ANOVA (or possibly MANOVA) rather than multiple $t$-tests, conventional rather than overly liberal significance levels, or the conservative Bonferroni correction to control the overall Type I error rate, produced findings that in some instances were of very doubtful statistical significance or replicability.

High-quality evaluations will be expensive. Child welfare researchers in Canada will thus have to aggressively pursue their fair share of the increased funding that has recently become available from federal granting bodies such as the Canadian Institutes for Health Research,
the Social Sciences and Humanities Research Council of Canada, or the National Centre for Crime Prevention. In addition, however, the provinces and territories will have to make a serious commitment to funding evaluations of their own child welfare programs, seeing this as an integral part of service delivery and an indispensable means of improving service effectiveness and efficiency. In this regard, it is encouraging that the government of at least one province, Ontario, has recently created a special fund for applied research and evaluation in child welfare. Administered by the new Research and Outcome Measurement Branch of the Ontario Ministry of Children and Youth Services, this fund is scheduled to grow in future. The same ministry has also recently decided to monitor outcomes on an ongoing basis of young people who have been in care for a year or more, by implementing, as of 2007, the developmentally oriented Looking After Children approach (Flynn, Ghazal, Legault, Vandermeulen, & Petrick, 2004) in all 53 Children’s Aid Societies in the province. Other provinces and territories are considering similar outcome-monitoring initiatives, which will complement but not replace the need for rigorous (i.e., preferably randomized) impact evaluations of specific programs. For, although outcome monitoring is useful in enabling decision-makers to assess outcomes and plan service improvements, unlike impact evaluation it does not allow changes in outcomes to be attributed to the program itself rather than to extraneous factors (Rossi et al., 2004).

Need for Evaluation of a Wider Range of Interventions

Of the 10 impact evaluations reviewed, 7 were directly or indirectly relevant to prevention, 5 to physical abuse or neglect, and 4 to sexual abuse. These topics seem to correspond reasonably well to the priorities formulated by the survey of stakeholders in child welfare (Vandermeulen et al., 2005), although interventions to prevent or treat sexual abuse (an issue of obvious importance) seem overrepresented, compared with those directed at the forms of maltreatment that have the highest incidence rates (Trocmé et al., 2005), namely, neglect, exposure to domestic violence, physical abuse, and emotional maltreatment.

Many other important evaluation topics suggest themselves, such as the effectiveness of foster care for young persons with conduct problems who are at high risk of serial eviction and subsequent severe deterioration in their mental health (Barber & Delfabbro, 2004); differential response, as a means of improving outcomes (Trocmé,
Knoke, & Roy, 2003); new strategies for adoption, especially of older children (Lindsey & Schwartz, 2004); promoting resilience to enhance life trajectories for young people in care (Flynn, Dudding, & Barber, 2006); or increasing the rate of post-secondary enrollments among young people in care by establishing Registered Educational Savings Plans.

Need for Implementation and Evaluation of Evidence-Based Practice Interventions

Four of the 10 impact evaluations reviewed were randomized controlled trials (RCTs). By and large, these more rigorous studies yielded information for practice that was considerably more clear-cut than that produced by the six quasi-experimental studies. This is no accident and is consistent with the tenets of evidence-based practice (EBP), a movement that Lindsey and Schwartz (2004) have termed one of the major advances in child welfare of the last 10 years and the fundamental requisite for improved practice. EBP was also identified as a research priority by the Canadian child welfare researchers and other stakeholders who responded to the electronic survey (Vandermeulen et al., 2005) mentioned at the outset of the present article.

In the context of child abuse services, Chaffin and Friedrich (2004) have defined evidence-based practice (EBP) as “the competent and high-fidelity implementation of practices that have been demonstrated safe and effective, usually in randomized controlled trials (RCTs)” (p. 1098). They also address the frequent situation where well-supported interventions (e.g., based on multiple RCTs carried out by independent research groups) are currently lacking:

In child maltreatment, very few intervention models meet the high standards required to designate a model as “well-supported.” EBP simply means favoring the best-supported available practices. Where well-supported or empirically validated treatments are available, they may be synonymous with EBP. Where there are no fully supported interventions, one must pick from among competing models with varying levels of support. For example, there might be one model tested in a single well-conducted randomized laboratory trial, along with a few quasi-experimental field studies or single-case multiple baseline studies, and with a robust clinical literature. This model might be compared to competing models having
no randomized trials and only anecdotal support. In this example, the first model might not meet full criteria as empirically validated, but compared to competing models, would clearly be the best supported and could be consistent with EBP. (p. 1104)

Chaffin and Friedrich (2004) offer several compelling arguments in favour of adopting, within child welfare, an EBP perspective, with its focus on services that objective and (as much as possible) randomized evaluations have shown to produce desired outcomes. According to these authors, EBP would replace the traditional culture of “evidence-informed” or “evidence-suggested” practice, which relies to a considerable degree not on outcome evidence but rather weak evaluation designs, subjective preferences, passing practice fashions, charismatic opinion leaders, advocacy, or political or social contexts. First, expert reviews of services in child abuse and neglect (e.g., Kauffman Best Practices Project, 2004; Saunders, Berliner, & Hanson, 2004) have concluded that “most field services provided to abused children and their families are not based on any clear evidence that the services actually work” (Chaffin & Friedrich, 2004, p. 1098). Second, even though clinical experience and single-group pre-post change designs may be informative about the process of service delivery, such data do not provide the comparison-based information needed to evaluate intervention effectiveness. Moreover, even somewhat stronger designs, such as non-randomized quasi-experimental studies, often have comparison-group deficiencies (as the present review found) and other weaknesses that generate less confident knowledge of intervention benefits than do RCTs. Third, EBP is particularly well suited to child welfare service systems for several reasons: child welfare systems aim at producing concrete, observable changes (e.g., ending child abuse, improving child safety, or increasing school success); as publicly funded systems, they are accountable for delivering or purchasing services that yield measurable outcomes; and, because they are composed of large-scale programs and organized teams of providers based on a common legislative mandate, they render the adoption of EBP interventions more feasible than if they were made up primarily of many independent, private-practice providers. Fourth, child welfare systems can already choose from an array of EBP interventions for improving outcomes for maltreated children and their families (see Chaffin & Friedrich, 2004, for the full references to the interventions that follow). EBP interventions in mental health include trauma-focused cognitive-behavioural therapy, abuse-focused cognitive-behavioural therapy, and parent-child interaction
therapy. An EBP for preventing physical abuse and neglect is the nurse-family partnership model (although the adaptation of this model by MacMillan et al. [2005] was not successful in preventing the recurrence of physical abuse or neglect). For preventing neglect, Project 12-Ways/SafeCare seems to be the best-supported EBP, at present. EBP interventions for reducing rates of physically abusive behaviour among abusive parents include parent-child interaction therapy and multisystemic therapy. An EBP intervention for sexually abused children is trauma-focused cognitive-behavioural therapy, as is multisystemic therapy for the treatment of juvenile sex offenders. For children in foster care with behavioural problems, a possible EBP model would be parent management training. For improving parenting, child welfare personnel could consider the triple-P positive parenting program or the incredible years parents, teachers and children training series, the most recent version of the Webster-Stratton (1989) model that Hughes and Gottlieb (2004) evaluated in their small-sample RCT.

Child welfare managers and evaluators in Canada would thus do well to investigate the EBP programs mentioned by Chaffin and Friedrich (2004). But these are not the only possibilities, and we conclude by mentioning two valuable compendia of additional EBP programs of mainly U.S. origin. (As our review suggests, evaluations of the impact of these interventions would be highly desirable, if the programs [Hughes & Gottlieb, 2004] or deliberate modifications of them [MacMillan et al., 2005] were to be implemented in the Canadian context.)

The U.S. Substance Abuse and Mental Health Services Administration (SAMHSA) has recently acquired the first compendium of EBP programs and placed it in the public domain for ready access. Known as the Communities That Care Prevention Strategies Guide (Channing-Bete Company, 2004), this source comprises over 50 “tested and effective” programs for youth problem prevention and positive development (see <http://www.channing-bete.com/positiveyouth/pages/CTC/prevention_strategies.html>). The second and more extensive compendium is the National Registry of Evidence-Based Programs and Practices (NREPP, which can be consulted at <http://modelprograms.samhsa.gov/template_cf.cfm?page=model_list>). The NREPP includes programs relevant to a wide array of domains (e.g., academic achievement, substance abuse, social and emotional competency, aggressive behaviour, etc.), in three categories. Promising programs are those that have been implemented and sufficiently evaluated and score at
least 3.3 on the 5-point NREPP scale on the parameters of integrity and utility. Effective programs have been well implemented and well evaluated, have been consistently positive in terms of their results, and score at least 4.0 on the 5-point NREPP scale on integrity and utility. Model programs have been well implemented and well evaluated, score at least 4.0 on the 5-point NREPP scale on integrity and utility, and have also agreed with SAMHSA to provide high-quality materials, training, and technical assistance for nationwide implementation. Canadian child welfare organizations should especially consider the NREPP model programs or effective programs, but even some of the promising programs may prove useful. As many as possible of these EBP interventions should undergo well-designed and adequately funded impact evaluations when implemented in Canada.

REFERENCES


Channing-Bete Company, Inc. (2004). Communities That Care Prevention Strategies Guide. South Deerfield, MA: Author. (This compendium of tested and effective programs was recently acquired and put in the public domain by the U.S. Substance Abuse and Mental Health Services Administration [SAMHSA]. For information on accessing it, see <http://www.channing-bete.com/positiveyouth/pages/CTC/prevention_strategies.html>.)


Robert J. Flynn is Professor of Psychology, Director of the Centre for Research on Community Services, and Principal Scientist at the Institute of Population Health at the University of Ottawa. He and his colleagues are currently evaluating applications in the Canadian context of two international initiatives: the Looking After Children approach to improving services for young people in out-of-home care, and the Communities That Care system for preventing youth problems and promoting positive youth development.

Danielle L. Bouchard is a full-time student in the Honours BA program of the School of Psychology and a former research assistant at the Centre for Research on Community Services at the University of Ottawa.