Family group decision making for children at risk of abuse and neglect

Aron Shlonsky, Kate Schumaker, Charlene Cook, David Crampton, Michael Saini, Elisabeth Backe-Hansen, Krystyna Kowalski

The Campbell Collaboration Social Welfare Group

This review is co-registered within both the Cochrane and Campbell Collaborations. A version of this review can also be found on Cochrane Library.
Protocol information

Authors
Aron Shlonsky¹, Kate Schumaker¹, Charlene Cook¹, David Crampton², Michael Saini¹, Elisabeth Backe-Hansen³, Krystyna Kowalski⁴

¹Factor-Inwentash Faculty of Social Work, University of Toronto, Toronto, Canada
²Mandel School of Applied Social Sciences, Case Western Reserve University, Cleveland, Ohio, USA
³NOVA Postboks 3223, Oslo, Norway
⁴SFI Campbell, The Danish National Centre for Social Research, Copenhagen, Denmark

Contact person
Aron Shlonsky
Factor-Inwentash Faculty of Social Work
University of Toronto
246 Bloor Street West
Toronto Ontario M5S 1A1
Canada

E-mail: aron.shlonsky@utoronto.ca

Dates

| **Assessed as Up-to-date:** |  |
| **Date of Search:** |  |
| **Next Stage Expected:** | 11 November 2009 |
| **Protocol First Published:** | Issue 3, 2009 |
| **Review First Published:** | Not specified |
# TABLE OF CONTENTS

**Protocol Information** .................................................................................................................. 2  
  Authors ................................................................................................................................. 2  
  Contact Person ..................................................................................................................... 2  
  Dates ........................................................................................................................................ 2  

**Background** ................................................................................................................................. 4  
  Description of the Condition ................................................................................................... 4  
  Description of the Intervention ............................................................................................... 5  
  How the Intervention Might Work ........................................................................................... 6  
  Why it is Important to Do this Review ..................................................................................... 6  

**Objectives** ................................................................................................................................... 7  

**Methods** ..................................................................................................................................... 7  
  Criteria for Considering Studies for this Review .................................................................... 7  
  Search Methods for Identification of Studies ........................................................................... 9  
  Data Collection and Analysis ..................................................................................................... 12  

**Acknowledgements** ....................................................................................................................... 17  

**Contribution of Authors** .............................................................................................................. 17  

**declarations of interest** ............................................................................................................... 17  

**Published Notes** ....................................................................................................................... 17  

**other references** ......................................................................................................................... 17  
  Additional References ............................................................................................................... 18  

**Sources of Support** ..................................................................................................................... 23  
  Internal Sources ......................................................................................................................... 23  
  External Sources ........................................................................................................................ 23  

**Appendices** .................................................................................................................................. 23  
  1 Screening Levels for Selecting Studies .................................................................................... 23  
  2 Systematic Information Retrieval Coding Sheet (SIRC) .......................................................... 25  
  3 Data Extraction Form .............................................................................................................. 25
BACKGROUND

Description of the condition

Freedom from abuse, neglect and violence is a basic right of all children (UNICEF 1989). However, child maltreatment remains a significant social problem that continues to affect many children worldwide, regardless of age, gender or culture (Pinherio 2006). Estimates of the incidence and prevalence of child abuse and neglect on an international scale are difficult to generate due to differences in how these phenomena are defined across cultures (Kempe 1978), the lack of epidemiological data from most regions other than North America, and the varying methodologies used for existing studies. Nonetheless, the World Health Organization (WHO) estimates that each year, millions of children around the world are victims and witnesses of physical, sexual and emotional violence (WHO 2006). Estimates of the incidence of child maltreatment in North America range from 18.67 child victims of maltreatment per thousand children in Canada (Trocmé 2005) to 41.9 out of 1,000 children in the US (Sedlack 1996). Children who experience abuse or neglect are at risk for a host of troubling short- and long-term psychosocial problems including anxiety, depression, social and behavioural problems, poor educational progress, and parenting difficulties (Clausen 2004; English 2005; Springer 2007).

In most high-income countries, primary responsibility for child protection rests with legislated child protection agencies, either operated or overseen by the state. While the specific mandate of these agencies differs across jurisdictions, statutory responsibilities usually include: receiving and investigating reports of suspected abuse and neglect; providing ongoing protection services to families in which children have been deemed at risk of future maltreatment; and the provision of out-of-home care when children cannot be safely cared for at home. While some families or children self-refer to a child protection agency, the majority of families referred are “involuntary” insofar as they have not asked for nor consented to the referral.

Traditionally, decision-making by child protection agencies has been professionally-driven, with workers conducting assessments of families’ problems and risk profiles, and determining a treatment plan with which families are asked to comply (Merkel-Holguin 2005; Rockhill 1999). This “top-down” approach to child protection services stems from the inherent power imbalance between child welfare professionals and their clients: poor families, single female-headed households, and visible minorities are disproportionately represented on child protection caseloads, while the workforce of child protection agencies is predominantly made up of middle class, educated, workers from the dominant culture (Pelton 1989; Dumbrill 2003). Additionally, the exclusion of parents from decision making is rooted in the presupposition that abusive or neglectful caregivers cannot make appropriate decisions concerning the care and protection of their children.

Within the last two decades there has been a growing acknowledgement of the need to work more collaboratively with families in the provision of child protection services (Littell 2000; Yatchmenoff 2005), to provide services that are sensitive to and protect children’s cultural identities (Connolly 2006), and to work with clients’ own definitions of problems and solutions rather than that of child protection professionals (Dumbrill 2003a; Lee 2004; Altman 2003). One response to these issues has been the introduction of Family Group Decision Making (FGDM) models as an alternative process in child protection.
Description of the intervention

Family Group Decision Making is an umbrella term [1] for practice models that shift planning for children away from “professionally driven” towards a more “family-centered” approach, with the premise that families are experts on their own situations, and as such, should be considered well qualified to contribute to plans designed to promote the safety and well-being of their children (Cunning 2006; Merkel-Holguin 2005). FGDM models typically involve one or more meetings between the extended family and other professionals, during which time a plan is developed for the care and safety of the child(ren) (Crampton 2004). Meetings may be held at several different points in the life of a case; for example, following the initial investigation, when it is determined that a child is at risk of coming into care, prior to re-unification, or at case closure.

There are numerous models of FGDM used internationally, including Family Group Conferencing (FGC), the Family Unity Meeting (FUM) model, Team Decision Making, Family Team Meetings, and Family Team Conferencing (CSSP 2002; Burford 2000). A central objective of all FGDM models is to provide the family with a stronger voice in decision-making than has typically been the case in traditional (often adversarial) child protection services. In this way, FGDM models hold the potential to address historical inequities in child protection, and to promote services that are in keeping with children and families’ cultural beliefs and identities (Merkel-Holguin 2005). Consensus does not exist in the field as to which of these and the numerous variants of these models best fit under the rubric of FGDM. The proliferation of different models primarily occurs in the United States, while most other countries draw upon the Family Group Conferencing model that originated in New Zealand. Therefore, that model deserves special attention here.

Family Group Conferencing (FGC) was created in New Zealand through a collaborative effort between Maori[1], governmental and community leaders, based on traditional Maori decision-making processes (AHA 2008; AHA 2008a; Connolly 1994). The impetus for the development of FGC stemmed from the overrepresentation of Maori children within both the child protection and juvenile justice systems, a growing acknowledgement that existing systems did not allow for participation of Maori families in decision-making about their children and did not fully recognize the inclusive view of “family” held by the Maori people[2] (Connolly 1994; Sundell 2001). The Family Group Conferencing model was introduced through the enactment of New Zealand’s Children, Young Persons and Their Families Act (1989), which replaced the former Children and Young Persons Act (1974); the addition of the word “family” to the title of the Act reflected a fundamental shift towards acknowledging the importance of the wider family system in the health and well-being of children, and in decision-making and planning for the future (Connolly 1994).

For the purposes of the current review, family group decision making interventions include all models that involve the convening of family and extended family members, identified friends and/or community members; child protection professionals; and, if needed or requested, other professionals, in an effort to collaboratively develop a plan to maintain child safety, facilitate stable and permanent living arrangements, and promote child and family well-being.

[1] The Maori are the indigenous people of Aotearoa New Zealand. They comprise approximately 14% of New Zealand’s population, living mostly on New Zealand’s North Island (Walker 2006).
The Maori concept of family differs from the traditional Western construct. The Maori words whanau, hapu and iwi, while not readily translatable, have a range of meanings from extended family to tribal affiliation, and comprise the familial structure upon which Maori society is based (Connolly 2004).

How the intervention might work

While the theory of how FGDM processes lead to specific client outcomes remains both understudied and underdeveloped (Crampton 2007), the basic underlying assumption of FGDM models is that, for a variety of reasons, solutions found within the family are likely to be more accepted and effective than those imposed by professionals (Sundell 2004). Most importantly, it is assumed that when families have a central role in planning, family members will be more likely to follow-through and maintain their involvement with the plan over the long term, leading to enhanced safety, permanence and well-being for children. Additionally, a primary aim of FGDM is to identify, seek out, and actively encourage the participation of the broad family and community system in caring for the children. In this way, FGDM models aim to strengthen the family and community network. Creating a strong network of support around the child and caregiver(s) may improve outcomes for children. In light of this objective, one of the aims of FGDM may also be to encourage the placement (voluntary, court-involved, temporary, longer term) of children who cannot be maintained with their birth parents.

Plans resulting from FGDM are also believed to be more consistent with the child and family’s cultural beliefs and identities (Merkel-Holguin 2005). One potentially promising outcome of FGC noted in the literature is the increased rate of placement with relatives when children do require out-of-home placement, and the greater likelihood that, when placed, children will remain with their siblings (Crampton 2007; Pennell 2006; Connolly 1999; Lupton 1999; Marsh 1998; Robertson 1996). Both anecdotal evidence and a growing body of research suggest that participants express consistently high satisfaction rates with the FGDM process (Sieppert 2000; Marsh 1998; Sundell 2004).

Finally, FGDM models frame families as competent and often explicitly focus on their strengths, with the aim of empowering families and shifting their experience of child protection service from one characterized by powerlessness to one of self-determination and collaboration (Lupton 1999). Literature across disciplines indicates that therapeutic settings that support clients’ sense of autonomy, relatedness and competence are more likely to bring about compliance with treatment, and greater transfer and maintenance of treatment gains (Deci 1985; Ryan 1995).

Why it is important to do this review

To date, FGDM models have been widely implemented in several countries, including New Zealand, the U.K., Canada, the United States, Australia, France, South Africa, Sweden, Norway, Denmark, Israel and the Netherlands (Cunning 2006; Faureholm 2005; Goldstein 2006; Cashmore 2000). While research exploring the effectiveness of these models in achieving key child and family outcomes is scant, some existing studies suggest that FGDM models may contribute to reducing future maltreatment for children (Pennell 2000; Marsh 1998), and increase the likelihood that children will remain within their extended family network when placed in out-of-home care (Crampton 2003; Crampton 2007a; Gunderson 2003). Other writers have reached more cautious conclusions, suggesting that
FGDM models have limited impact on safety outcomes for children when compared to regular child protection services (Sundell 2001; CSSR 2004). Several researchers have noted the limitations of existing research, and call for studies that involve larger sample sizes and more rigorous control group designs (Lupton 1999; Crampton 2007), and research that focuses on testing which specific elements of FGDM might lead to its effectiveness (Crampton 2004). There have been few longitudinal studies into the effectiveness of FGDM models, and no systematic review has been conducted to synthesize existing research.

Family Group Decision Making models are conceptually compelling and consistent with social work values and principles of empowerment and culturally appropriate practice. These models have been widely implemented internationally in child welfare contexts. In addition to the precedent-setting New Zealand legislation, several countries and jurisdictions have legislation or policies encouraging the use of FGC or FGDM in cases of child abuse and neglect, and provide government funding for FGDM programs.

Despite the widespread support and investment in FGDM, key outcomes for children and families who receive FGDM interventions (safety, permanence and well-being) are not well documented, particularly over the longer term (Connolly 1994; Lupton 1999; Maluccio 2000; Sundell 2004; Connolly 2004), and no review has systematically synthesized existing research. The combination of the widespread popularity of FGDM along with the limited (and mixed) evidence of its effectiveness makes this review particularly important. A review will begin to clarify our understanding of the extent to which FGDM models successfully prevent future maltreatment and facilitate permanent, stable living situations for children. This information needs to be readily available to communities where FGDM models have already been implemented or are under consideration.

**OBJECTIVES**

To assess the effectiveness of the formal use of FGDM in terms of child safety, permanence (of child's living situation), child and family well-being, and client satisfaction with the decision-making process.

**METHODS**

Criteria for considering studies for this review

Types of studies

Studies will be eligible for this review if they: 1) used random assignment to create treatment and comparison or control groups; or 2) used parallel cohort designs in which groups were assessed at the same points in time (i.e., quasi-experimental designs that include groups assessed at the same time as opposed to a historical cohort). Single-group designs and single-subject designs will be excluded (see 'risk of bias' section for further details on included designs).

Types of participants

Children and young people aged 0-18 years who have been the subject of a child maltreatment investigation.
Types of interventions

Any form of Family Group Decision-Making (FGDM) used in the course of a child maltreatment investigation or during the course of services arising from such an investigation.

This involves convening family, extended family, identified friends and/or community members along with child protection professionals and other professional, community-based collaterals in an effort to collaboratively develop a plan to maintain child safety, facilitate stable and permanent living arrangements, and promote child well-being. Therefore, studies will be included in the review if they involve: 1) a concerted effort to convene family, including extended family, friends and community members; and 2) child protection professionals (as well as other professional service providers) participating in; 3) a planned meeting with the intention of working collaboratively to develop a plan for the safety, permanence and well-being of child(ren); and 4) with a focus on family-centred decision-making.

Types of outcome measures

Primary outcomes

Prevention of child maltreatment and increasing family permanence and Placement Stability. Prevention of child maltreatment is measured by the following constructs, and these are arrayed in order of 'best' (i.e., most likely to be accurate) to 'worst' (less likely to be accurate):

- Substantiated/verified/indicated referrals to a child protection authority (gold standard)
- Referrals (not verified) to a child protection authority
- Parent self-report
- Child self-report
- Collateral party report

Analyses may include all of these types of measurements, but these will first be grouped by indicator, 'best' source will be preferred, and studies will be analyzed separately prior to synthesis.

Family Permanence and Placement Stability will be measured by the following indicators:

- For children residing in the homes of their birth parents, entry into foster care or other out-of-home placement will be interpreted in a two different ways. First, any and all placements will be analyzed as a negative outcome (i.e., FGDM, in this instance, is being used to prevent any placement, including kinship placements, into out-of-home care). However, kinship care (placement in out-of-home care with relatives) is a unique type of care that FGDM may actually facilitate as an alternative to placement to other forms of out of home care. Therefore, Kinship care will also be interpreted as a positive outcome and analyzed separately (i.e., does the rate of placement into kinship care differ between children receiving FGDM versus children who do not), and the interpretation of this set of analyses
will be made within the context of FGDM's stated purpose (i.e., involving family members in child placement decisions).

- For children residing in out-of-home care (i.e., foster care, kinship care, group care), legal permanence will be interpreted as a positive outcome (i.e., reunification with birth parents, adoption by related or non-related caregivers, placement with relative caregivers, legal guardianship / legal custody by related or non-related caregivers (i.e., FGDM, in this instance, is being used to facilitate family permanence).
- For children residing in out-of-home care, long-term foster or kinship care arrangements with the same caregiver (i.e., placement stability) will also be considered positive outcomes, though such outcomes are generally held in low regard when compared to legally permanent homes.

Since the primary outcomes listed here are generally events rather than perceptions or subjective impressions, the gold standard or 'best' indicator for the measurement of primary outcomes will be official reports found in administrative data and case files. For prevention of child maltreatment, there may be some studies that utilize self-report or a report from a secondary party. In such cases, only studies using standardized tools measuring the occurrence of child maltreatment and/or family violence, for example the Conflict Tactics Scale (*Straus 1996*), will be included.

Studies will only be included in the analysis of primary outcomes if subjects are followed for at least six months after the intervention to allow for sufficient time to observe outcomes. For included studies, longest common follow-up will be used but separate analysis of short-term (e.g., 6 months) and long-term (e.g., 3+ years) outcomes will be conducted for each primary outcome measure.

**Secondary outcomes**

Secondary outcomes include child **well-being**, and **client satisfaction** with the FGDM process and plan.

**Well-being** will be measured through the inclusion of studies that employ commonly used instruments for measuring child well-being such as the Child Behaviour Checklist (*Achenbach 1991*). Instruments will be coded for information on reliability and validity, and these will be analyzed to ascertain whether they covary with respect to study outcomes.

**Client satisfaction** will be measured by any tool developed for the FGDM context, such as the “Amount and Adequacy of Say Scale” and the “Decision Process Ranking Scale” (*Pennell 1995*). Instruments will be coded for information on reliability and validity, and these will be analyzed to ascertain whether they covary with respect to study outcomes.

**Search methods for identification of studies**

Both published and unpublished work will be considered eligible for the review. A Trial Search Coordinator (Carmen Logie) will be responsible for coordinating this activity. To the greatest extent possible, the search will not be restricted to any single language or nationality.

A Systematic Information Retrieval Coding Sheet (SIRC) has been developed (see Appendix 2) to record each search for the review. The SIRC will be used to log results for each database and grey literature searched and will include:
1) The date(s) of the search;
2) The name of the researcher;
3) The database used for the search;
4) The specific search terms used in combination (including limiters and expanders); and
5) The number of results for each search strategy.

Such recording facilitates replication of the search strategy. Furthermore, the search strategy will be saved and “copied and pasted” into the review to avoid editing errors.

Electronic searches
Electronic searches for the identification of appropriate studies will include the following bibliographic Databases:

1. Cochrane Central Register of Controlled Trials (CENTRAL)
2. MEDLINE
3. The Campbell Collaboration Register of Controlled Trials (C2-SPECTR)
4. PsycINFO
5. EMBASE
6. Database of Abstracts of Reviews of Effects (DARE)
7. ASSIA (applied social sciences)
8. ERIC
9. CINAHL
10. International Bibliography of the Social Sciences
11. Caredata (social work)
12. Social Work Abstracts
13. Social Sciences Abstracts
14. Child Abuse and Neglect (CANDIS)
15. Australian Family and Society Abstracts Database
16. Dissertation Abstracts International (DAI)

To ensure maximum sensitivity and specificity, subject headings and word text will be searched in a systematic process.

The search that will be used for Medline is as follows and will be modified accordingly for the other databases listed:

1. family group.tw.
2. family decision.tw.
3. family decisionmaking.tw.
4. family conferenc$.tw.
5. family unity.tw.
6. family team.tw.
Searching other resources

Reference lists
Reviewers will check the reference lists of all relevant articles that are obtained, including those from previously published reviews. Potentially relevant articles that are identified will be retrieved and assessed for possible inclusion in the review.

Personal communication
Face-to-face discussions at meetings, emails, requests on list-servs, and formal letters of request for information from authors, presenters and experts will be solicited to assist the review team to locate relevant studies. A list of the inclusion criteria for the review, along with a sample of relevant articles, will be sent to these key informants along with the request for studies. The list of experts to be contacted will include principle investigators of eligible studies, program developers, and authors of previous reviews of relevant literature.

Handsearching journals
Journals relevant to child maltreatment will be handsearched by trained researchers to uncover relevant studies not found by electronic database searches. In addition, trained reviewers will search reference lists of relevant articles. These include:

1) Child Welfare
2) Children and Youth Services Review
3) Social Service Review
4) Child Maltreatment
5) Child Abuse and Neglect
6) Journal of Social Services Research
7) Social Work
8) Research on Social Work Practice  
9) Social Work Research  
10) Child Abuse Review  

**Grey Literature**  
Special attention will be made to search and collect relevant studies captured in the grey literature. Specifically, the review will include the following strategies to locate articles:  
1) Conference Proceedings (e.g. PapersFirst and ProceedingsFirst, both accessed through the University of Toronto library system); 2) Research Reports (e.g. http://www.evaluationcanada.ca/site.cgi?s=6&ss=8); 3) Government Reports and Policy Documents (e.g. http://www.gc.ca/publications/publication-eng.html, http://www.usa.gov/, http://publications.gov.au/, http://www.natlib.govt.nz/collections/types-of-items/government-publications, and http://europa.eu/index_en.htm); 4) Book Chapters; 5) Dissertations (e.g. ProQuest Dissertations and Theses, Theses Canada Portal http://www.collectionscanada.gc.ca/thesescanada/index-e.html); 6) Personal Networks; and 7) Research Organizations’ Web Sites. Grey Literature web-based sites will be searched to uncover this unpublished literature, such as Grey.Net (http://www.greynet.org/index.html)  

**Data collection and analysis**  

**Selection of studies**  
Titles and abstracts of studies yielded by the searches will be independently screened by two reviewers to determine their eligibility for inclusion in the review. The screening of the studies will be carried out by a three-stage procedure (see Appendix 1) where each screening point will be established in ‘level’ format and each level will consist of increasing scrutiny of the studies based on the inclusion and exclusion criteria of the review.  

**Data extraction and management**  
Two review authors, using a data extraction form, will independently extract data on participants, methods, interventions, outcomes, and results (Appendix 3). We will obtain missing data from study authors if possible.  
Details to be extracted will include:  
1) Study: information regarding the author(s); year of publication; source; country; and language  
2) Characteristics of Setting and Participants: eligibility criteria for participants; explanation of recruitment procedures, setting (country, location, clinical/non-clinical); demographic features of the sample  
3) Sampling: sample sizes for treatment and control; whether power analysis was used to determine sample size; allocation to the treatment and control; explanation of method used to generate the allocation  
4) Research Design: Type of design including major features such as random selection, random assignment, and non-equivalent control group. Features will be assessed according to ‘Assessment of risk of bias’ categories as described below
5) Intervention Data: nature of interventions (for treatment and comparison/control groups); FGC, FUM, or some other form of Family Group Decision-Making; aim of intervention; length of intervention, whether manuals were used, whether fidelity checks were included, information on possible contamination reported.

6) Outcome Data: primary and secondary outcomes, measures used, information on reliability/validity of measures.

7) Results: attrition at post intervention and follow-up; number excluded from the analysis; length of follow-up; statistical methods; type of data effect size is based on; data needed for effect size calculations.

Assessment of risk of bias in included studies
Risk of bias will be assessed independently by two review authors according to the Cochrane Collaboration Handbook (Higgins 2008). Review authors will independently assess the risk of bias within each included study based on the following six domains with ratings of 'Yes' (low risk of bias); 'No' (high risk of bias) and 'Unclear' (uncertain risk of bias).

Reporting bias within each of the included studies (e.g. selection, measurement, attrition bias) will be reported in the results and discussion.

Sequence generation
The method used to generate the allocation sequence is described in detail so as to assess whether it should have produced comparable groups.

Review authors' judgment: was the allocation concealment sequence adequately generated?
Ratings: 'Yes' (low risk of bias); 'No' (high risk of bias) and 'Unclear' (uncertain risk of bias)

Allocation concealment
The method used to conceal allocation sequence is described in sufficient detail to assess whether intervention schedules could have been foreseen in advance of, or during, recruitment; review authors' judgment: was allocation adequately concealed?
Ratings: 'Yes' (low risk of bias); 'No' (high risk of bias) and 'Unclear' (uncertain risk of bias)

Blinding
Any measures used to blind participants, personnel and outcome assessors are described so as to assess knowledge of any group as to which intervention a given participant might have received.

Review authors' judgment: was knowledge of the allocated intervention adequately prevented during the study?
Ratings: 'Yes' (low risk of bias); 'No' (high risk of bias) and 'Unclear' (uncertain risk of bias)

Incomplete outcome data
If studies do not report intention-to-treat analyses, attempts are made to obtain missing data by contacting the study authors. Data on attrition and exclusions are extracted and reported, as well as the numbers involved at measurement period (compared with total
randomized at pre-test), whether the reasons for attrition/exclusion are reported or obtained from study authors, and whether the study authors perform any re-inclusions of missing data in analyses.

Review authors' judgment: were incomplete data dealt with adequately by the reviewers? Ratings: 'Yes' (low risk of bias); 'No' (high risk of bias) and 'Unclear' (uncertain risk of bias)

**Other sources of bias**

Was the study apparently free of other problems that could put it at a high risk of bias (These will be determined once the included studies are considered, but will likely include contamination as this is present in at least one known study)

Ratings: 'Yes' (low risk of bias); 'No' (high risk of bias) and 'Unclear' (uncertain risk of bias)

**Measures of treatment effect**

For binary outcome data, effect sizes will be calculated as odds ratios (OR) with 95% confidence intervals. Continuous data will be converted into standardized mean differences (SMDs) and presented with 95% confidence intervals. When necessary, we will use formulas suggested by Lipsey and Wilson (2001) to convert correlation coefficients, F ratios, t-values, and chi-square values into SMDs. Hedges' $g$ will be used to correct for small sample bias.

We assume that there will be unexplained sources of heterogeneity across studies; hence assumptions of the fixed effect model (that all studies provide estimates of a single population effect size) are likely to be untenable. Therefore, the random effects model will, in all likelihood, be used for pooling results. Results for randomized experiments and quasi-experimental designs will be pooled and reported separately. As well, results for conceptually-distinct outcomes will be reported separately. If a study reports two separate measures of the same outcome, these will be averaged and the newly created effect size for the same outcome will represent the study in the analysis. Where possible, we will also explore differences between models (e.g., FGC and FUM).

Aron - Odds Ratio Statement

**Unit of analysis issues**

**(a) Cluster-randomised studies**

Where trials have used clustered randomization, we anticipate that study investigators would have presented their results after appropriately controlling for clustering effects (robust standard errors or hierarchical linear models). If it is unclear whether a cluster-randomized trial has used appropriate controls for clustering, the study investigators will be contacted for further information. Where appropriate controls were not used, individual participant data will be requested and re-analysed using multilevel models which control for clustering. Following this, effect sizes and standard errors will be meta-analysed in RevMan using the generic inverse method (Higgins 2008). If appropriate controls were not used and individual participant data are not available, statistical guidance will be sought from the Cochrane Methods Group and external experts as to which method to apply to the published results in attempt to control for clustering. If there is insufficient information to control for clustering, outcome data will be entered into RevMan using individuals as the
units of analysis, and then sensitivity analysis will be used to assess the potential biasing effects of inadequately controlled clustered trials (Donner 2001).

(b) Cross-over studies
Due to the nature of the intervention (Family Group Decision Making) we do not anticipate cross-over studies will be identified.

Dealing with missing data
In cases where data are missing (e.g., subgroup means and standard deviations, valid Ns), we will contact the author(s) of the primary studies and try to obtain missing information. We will also search for unpublished reports or other write-ups of the studies. If an author is unable or unwilling to provide information, we will exclude the study if there is inadequate information to proceed. Data on excluded subgroups (e.g., program drop-outs) will be sought as well and Intent to Treat (ITT) analysis will be conducted wherever possible. Studies using ITT or where an ITT analysis can be conducted will be compared with studies where this information is unavailable.

Assessment of heterogeneity
Statistical heterogeneity in the outcome measures will be assessed using the Q-statistic and the associated p-value for each analysis and the I² statistic (Higgins 2008). The I² statistic will determine the percentage of variability that is due to heterogeneity where a value greater than 50% suggests moderate heterogeneity.

Assessment of reporting biases
Publication and small sample bias will be assessed with graphical inspection of funnel plots, and "trim and fill" methods that estimate treatment effect by adjusting for the number and outcomes of missing studies (Sutton 2000).

The authors will deal with selective outcome reporting by searching for the original reports upon which many of the included studies will have been based and comparing the types of outcomes included with the outcomes reported in the published studies. The authors will also search conference abstracts and other sources of gray literature for earlier or additional studies and compare the range of outcomes. In cases where selective outcome reporting is suspected, the study author will be contacted. The assessment of risk of bias due to selective reporting of outcome will be made for each study as a whole.

Data synthesis
Data synthesis will be conducted using RevMan 5 and Comprehensive Meta-analysis 2.0. The determination of independent findings will be completed using the following procedures: First, studies may have included more than one measure of the outcome; therefore, to ensure statistical independence of study findings, each measure of the outcome will be analyzed separately.

We will do separate analyses for absolute vs. relative effects by doing separate analyses for studies that use no-treatment or wait-list controls and for studies that compare two different treatments (Absolute effects will be compared with no treatment, e.g., wait lists; relative effects will be compared to other treatments, e.g., regular child welfare service provision or TAU).
Multiple outcomes for dependent or overlapping samples (i.e., multiple treatments compared against one control group) will be coded separately. We will select only one effect size for inclusion in the meta-analysis based upon conceptual relevance (i.e., FGDM versus another commonly used intervention), sample size, and completeness of information. For studies that include multiple follow-ups, these will be divided into separate intervals (i.e., effects within 6 months, 7-12, 13-24, more than 24 months) and we will do separate meta-analyses for each separate interval.

If there is substantial and unreconcilable heterogeneity with respect to intervention definition, outcomes specification, or measures used, there may come a point at which a meta-analysis becomes untenable (e.g., the population, intervention, measures, or outcomes differ so substantially that combining studies would make little sense), we will conduct a narrative synthesis of the studies. The narrative synthesis would still address each outcome and would detail the studies included, their methodological strengths and weaknesses (as evaluated using CONSORT guidelines), and, based on these, a conservative appraisal of the merit of the intervention will be provided. This is not the preferred course since such a synthesis would not produce any clear recommendations for practice and policy. Nonetheless, it would provide readers with an honest appraisal of systematically gathered studies rather than a more biased literature review, and might prompt further rigorous studies of this commonly used intervention.

Syntheses of higher quality studies are often considered more accurate than syntheses of lesser quality studies. However, there is the possibility that less rigorous designs may produce less biased effect sizes (i.e., they may measure outcomes better or use less biased samples). Moderator analysis (if possible) will be used to control for some of these differences. However, if different trends emerge based on study design (e.g., RCT's v. non-equivalent control group designs), and these can not be controlled for in the analysis, results will be presented separately and the possible reasons for such differences will be discussed.

Subgroup analysis and investigation of heterogeneity

To the greatest extent possible, methodological and clinical heterogeneity among studies will be explored in terms of variations associated with overall study design (experimental and quasi-experimental designs), baseline characteristics (i.e., child(ren) in care v. child(ren) in-home), type of FGDM (e.g., FGC v. FUM), comparison condition (variations in TAU if found), duration of follow-up, and outcome measures (i.e., different measures of a single outcome). If sufficient studies are found, we will also examine effects of interventions with different subpopulations (e.g., type of maltreatment, type of out-of-home care provider). Moderator analysis will be performed using the ANOVA analog (for categorical moderators) and/or meta-regression (for continuous moderators).

Sensitivity analysis

Sensitivity analysis will be performed to assess the robustness of conclusions to quality of data and approaches to analysis (see risk of bias section). Sensitivity analysis will be performed by reanalysis, excluding studies with poor quality indicators (e.g., high attrition, differential attrition, lack of intent-to-treat analysis, lack of controls for baseline differences).
ACKNOWLEDGEMENTS

The following people graciously responded to requests for information: Lisa-Merkel Holguin, American Humane Association; Ted Keys, Oregon Department of Human Services; Angela Rodgers, Portland State University. We would also like to acknowledge the early contribution of Tony Newman, who originally registered this title. Anne-Marie Jørgensen, SFI Campbell for searches of the Nordic data bases.

CONTRIBUTION OF AUTHORS

Aron Shlonsky contributes substantive child welfare content and systematic review expertise. He is the lead author and will guide all facets of the review including write-up, review of articles, and synthesis.

Kate Schumaker contributes substantive content expertise and is responsible for a substantial portion of the write-up, review of studies, and some synthesis.

David Crampton contributes substantive North American content expertise, locating English language grey literature, review of studies, and interpretation of results.

Charlene Cook and Michael Saini contribute methodological expertise, including meta-analysis.

Elisabeth Backe-Hansen contributes searches of the Nordic published and European grey literature, review of studies, as well as substantive European content expertise.

Krystyna Kowalski contributes searches of the Nordic literature, hand searching of English journals, and review of studies.

DECLARATIONS OF INTEREST

Aron Shlonsky, Katherine Schumaker, Michael Saini, Charlene Cook, Krystyna Kowalski and Elisabeth Backe-Hansen have no declarations of interest to report.

David Crampton currently receives funding from the Annie E. Casey Foundation to support his participation in an evaluation of Team Decisionmaking (TDM) and other child welfare reforms associated with the foundation’s Family to Family Initiative. The Casey Foundation promotes TDM. He previously conducted an evaluation of a Family Group Decision Making program in Kent County in Michigan which was supported by the Grand Rapids Community Foundation, W. K. Kellogg Foundation and the United States Children’s Bureau. He has written nine articles about FGDM that were published by the American Humane Association (AHA); AHA promotes FGDM. He has authored or co-authored seven articles about FGDM and/or TDM in peer-reviewed publications. He also wrote a book chapter about the use of FGDM for older youth in foster care with Joan Pennell for which they received support from Casey Family Services.

PUBLISHED NOTES

This protocol is co-registered within the Cochrane Collaboration.

OTHER REFERENCES
Additional references

Achenbach 1991

AHA 2008
American Humane Association. Historical timeline of FGDM. American Humane Association Website Association:

AHA 2008a

Altman 2003

Burford 2000

Cashmore 2000

Clausen 2004

Connolly 1994

Connolly 1999

Connolly 2004
Connolly M. A perspective on the origins of Family Group Conferencing. American Humane FGDM Issues in Brief

Connolly 2006

Crampton 2003

Crampton 2004

Crampton 2007

Crampton 2007a

CSSP 2002

CSSR 2004

Cunning 2006

Deci 1985

Donner 2001

Dumbrill 2003

Dumbrill 2003a

English 2005

Faureholm 2005

Goldstein 2006

Gunderson 2003

Higgins 2008

Kempe 1978

Lee 2004

Littell 2000

Lupton 1999
Marsh 1998

Merkel-Holguin 2005

Pelton 1989

Pennell 1995

Pennell 2000

Pennell 2006

Pinheiro 2006

Robertson 1996

Rockhill 1999

Ryan 1995

Sedlack 1996
Sieppert 2000

Springer 2007

Straus 1996

Sundell 2001

Sundell 2004

Sutton 2000

Trocme 2005

UNICEF 1989

Walker 2006

WHO 2006

Yatchmenoff 2005
SOURCES OF SUPPORT

Internal sources

• Factor-Inwentash Child Of Social Welfare, University of Toronto, Canada
• SFI Campbell, Nordic Campbell Centre, Denmark

External sources

• Splane Family Fund, Canada

APPENDICES

1 Screening Levels for Selecting Studies

1) Initial Screening (level 1)
The first stage will consist of an initial screening to quickly determine whether a study might be appropriate for the review based on the study’s title and abstract. If there is not enough information in the title and abstract to make such decisions, then full text articles will be retrieved. Any disagreements will be resolved by retrieving the full text. The purpose of this initial screen will be to include all possible relevant studies related to the objectives of the systematic review and the inclusion and exclusion criteria. Level one screening will consist of the following questions:

• Does the population consist of children and youth who are or have been investigated by the child protection system for child maltreatment? Yes / No
• Is there an intervention related to family group conferencing experienced by the population included in the study? Yes / No

2) Strict Screening (level 2)
The second stage will consist of a strict screening process where two reviewers will independently review full copies of articles to determine whether studies should remain in the review based on the inclusion and exclusion criteria. Specific reasons for exclusion at this stage will be documented for each study. Any disagreements will be resolved by a third reviewer. Level two screening will consist of a double check of the level one screening as well as the following questions:

• Did the study evaluate the intervention administered to children and youth, age 0-18? Yes / No
• Did the evaluation use an experimental or parallel cohort research design? Yes / No
• Is there a parallel cohort (comparison or control group)? Yes / No

3) Data extraction and management (level 3)
The third stage uses a data extraction form to record data from the articles that have made it past the two previous screenings. The study details will be extracted, using a data extraction sheet, by two independent reviewers (see Appendix 2). Differences between coders will be identified and resolved to ensure consistent extraction and management of the data and to establish inter-rater reliability. Any discrepancies will subsequently be resolved by referral back to the source of the material and conflicts will be resolved by a third reviewer based on the original source. If necessary, we will seek additional information from the original investigators.
2 Systematic Information Retrieval Coding Sheet (SIRC)

Project: __________________________________________________________________________

Reviewer: __________________________________________________________________________

Date(s) of Search: ____________________________________________________________________

Search Method:  
Electronic Database: Name: _____________________
GREY LITERATURE: Name: _____________________
Other: Name: _____________________

Language(s): ________________________________________________________________________

Date Range: ________________________________________________________________________

Description of Search: __________________________________________________________________

Search Terms

<table>
<thead>
<tr>
<th>Population</th>
<th>Intervention</th>
<th>Outcome</th>
<th>MOLES</th>
</tr>
</thead>
</table>

Search Term Combinations (including all limiters and expanders) | Results
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 Data Extraction Form

Study Level
<table>
<thead>
<tr>
<th><strong>Administration</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reference Number (Study identifier):</strong></td>
</tr>
</tbody>
</table>

*If multiple documents were used to code this study, indicate the supplemental study ID numbers*

| **Cross reference document identifier:** |

| **Reviewer:** |

| **Date(s) of the Review:** |

<table>
<thead>
<tr>
<th><strong>Source</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s):</strong></td>
</tr>
</tbody>
</table>

| **Year of Publication:** |

| **Title:** |

| **Source:** |
| [ ] Book |
| [ ] Conference Paper |
| [ ] Peer Review Journal Article |
| [ ] Non-Peer Review Journal Article |
| [ ] Dissertation |
| [ ] Report |
| [ ] Government Publication |
| [ ] Other: _________________________________ |

| **Search Method:** |
| [ ] Electronic search: |
| [ ] Hand search: |
| [ ] Grey Literature: |
| [ ] Recommendation: |
| [ ] Other: |

| Number of different “modules included in the report” | ________ |
Is the same control/comparison group used in different modules (1 = Yes; 0 = No)

**Treatment - Comparison Contrast Level**

**Program Description**

Program Description:

---

**Primary Treatment Type:**

Family Group Decision-Making includes:

1. Family Group Conferencing (FGC)
2. Family Unity Meeting Model (FUM)
3. Team Decision Making
4. Family Team Conferences
5. Family Team meetings

**In what format or social setting is the treatment delivered:**

1. One-on-one
2. Group setting
3. Family setting
4. Internet-based
5. Mixed (any combination of the above)
6. Unclear

**Who delivers the treatment?**

1. Child Welfare Worker
2. Community Worker
3. Third-party
4. Nonprofessional
5. Other

**Length of treatment type in months:**

a. Minimum [ ]
b. Maximum [ ]
c. Mean [ ]
d. Fixed (same for all subjects) [ ]

Length of follow-up program component (in weeks) [ ]

**Details of the intended treatment type included:**
1 = Yes; 0= No

**Details on the implementation of the treatment type included:**
1 = Yes; 0= No

**Manuals used for implementation of the treatment type:**
1 = Yes; 0= No

**Fidelity checklist used for the implementation of the treatment type:**
1 = Yes; 0= No

**Information on possible contamination reported**
1 = Yes; 0= No

**Describe the program for the comparison group if other than no treatment or treatment as usual:**

**What happens to the comparison group?**

[ ] No treatment
[ ] Waiting list (treatment begins at post)
[ ] Waiting list (treatment begins at follow-up)
[ ] Waiting list (treatment begins after study)
[ ] Minimal treatment
[ ] Alternate treatment

Where is the comparison drawn from:

**Methodological Rigor**

Use of control variables in statistical analyses to account for initial group differences (1 = Yes; 0= No)

Use of random assignment to conditions (1= Yes; 0= No)
If not random assignment, use of subject level matching (1= Yes; 0= No)

Matching variable(s) appropriate (1= Yes; 0=No)

Prior measurement of child maltreatment of family permanence (1= Yes; 0= No)

Prior measurement of child well-being or client satisfaction (1= Yes; 0= No)

Measurement of prior risky online behaviors (1= Yes; 0= No)

Rating of initial group similarity (7 = highly similar; 1= dissimilar)

Anchors:  7  Randomized design large N or small N with matching
5  Nonrandomized design with strong evidence of initial equivalence
1  Nonrandomized design, comparison group highly likely to be different or known different that are related
to future cyber abuse.
Was attrition discussed in the study reported? (1= Yes; 0= No)

---

### Sample Level Coding Sheet

<table>
<thead>
<tr>
<th>Characteristics of Setting and Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Description:</td>
</tr>
<tr>
<td>Sample description treatment group:</td>
</tr>
<tr>
<td>Sample description comparison group</td>
</tr>
<tr>
<td>Explanation of recruitment procedures:</td>
</tr>
<tr>
<td>Are the subjects included in the study clearly defined in terms of demographic features</td>
</tr>
<tr>
<td>(age, sex, ethnicity, presence/absence of condition for eligibility criteria)?</td>
</tr>
<tr>
<td><strong>Population Characteristics:</strong></td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td><strong>Comment:</strong></td>
</tr>
<tr>
<td><strong>Sampling</strong></td>
</tr>
<tr>
<td><strong>Total number of individuals at beginning of the study:</strong></td>
</tr>
<tr>
<td>Treatment Group N=</td>
</tr>
<tr>
<td>Control Group N=</td>
</tr>
<tr>
<td>Total Sample N=</td>
</tr>
<tr>
<td><strong>Use of power analysis to determine sample size:</strong></td>
</tr>
<tr>
<td>Yes [ ] No [ ] Not Clear [ ]</td>
</tr>
<tr>
<td><strong>Outcome Level Code Sheet</strong></td>
</tr>
<tr>
<td><strong>Outcome Data</strong></td>
</tr>
<tr>
<td><strong>Outcome indicator for Family Group Decision Making:</strong></td>
</tr>
<tr>
<td>1 child maltreatment;</td>
</tr>
<tr>
<td>2 family permanence</td>
</tr>
<tr>
<td>3 child well-being</td>
</tr>
<tr>
<td>4 client satisfaction</td>
</tr>
<tr>
<td><strong>Outcome measures relevant to goals of intervention</strong></td>
</tr>
<tr>
<td>Yes [ ] No [ ] Not Clear [ ]</td>
</tr>
<tr>
<td><strong>Explanation of measurement instrument and information regarding reliability and validity</strong></td>
</tr>
<tr>
<td>Yes [ ] No [ ] Not Clear [ ]</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
</tr>
<tr>
<td>Outcome:</td>
</tr>
<tr>
<td>Instrument:</td>
</tr>
</tbody>
</table>
Type of measurement scale
(1=Dichotomy; 2= Tricotomy; 3= 4-9 discrete ordinal categories; 4= >9 discrete ordinal categories or continuous)

Source of data
(1=self-report; 2= other report (teacher, parent), 3= official report, 4= other, 5=unclear)

Is this a valid and reasonable measure of reduction of cyber abuse? (1 = questionable; 2= acceptable)

---

**Effect Size Level Code Sheet**

**Data Reported**

<table>
<thead>
<tr>
<th>Identifying Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study identifier [ ]</td>
</tr>
<tr>
<td>Module identifier [ ]</td>
</tr>
<tr>
<td>Sample identifier [ ]</td>
</tr>
<tr>
<td>Outcome identifier [ ]</td>
</tr>
<tr>
<td>Effect size identifier (number each effect size within a study sequentially) [ ]</td>
</tr>
</tbody>
</table>

**Pages where data are found**

---

**Effect Size Information**

Effect size type
1. Baseline (pretest; prior to start of intervention)
2. Post-test (first measurement point post intervention)
3. Follow-up (all subsequent measurement points, post intervention)

Time frame in months captured by measure
a. Minimum [ ]
b. Maximum [ ]
c. Mean [ ]
d. Fixed [ ]

**Effect Size Data**
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment group sample size</td>
<td>[ ]</td>
</tr>
<tr>
<td>Comparison group sample size</td>
<td>[ ]</td>
</tr>
<tr>
<td>Treatment group mean (decimal)</td>
<td>[ ]</td>
</tr>
<tr>
<td>Comparison group mean (decimal)</td>
<td>[ ]</td>
</tr>
<tr>
<td>Adjusted? (1=Yes; 0=No)</td>
<td></td>
</tr>
<tr>
<td>Treatment group standard</td>
<td>[ ]</td>
</tr>
<tr>
<td>Comparison group standard</td>
<td>[ ]</td>
</tr>
<tr>
<td>Treatment group successful</td>
<td>[ ]</td>
</tr>
<tr>
<td>Comparison group successful</td>
<td>[ ]</td>
</tr>
<tr>
<td>t-value from independent t-test</td>
<td>[ ]</td>
</tr>
<tr>
<td>Exact p-value</td>
<td>[ ]</td>
</tr>
<tr>
<td>Chi-square value</td>
<td>[ ]</td>
</tr>
<tr>
<td>Correlation coefficient (point)</td>
<td>[ ]</td>
</tr>
<tr>
<td>Correlation coefficient (phi)</td>
<td>[ ]</td>
</tr>
<tr>
<td>Computer Calculated ES</td>
<td>[ ]</td>
</tr>
<tr>
<td>Hand Calculated ES</td>
<td>[ ]</td>
</tr>
<tr>
<td>Hand Calculated SE of ES</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

**Additional Comments**