Family group conferencing for children and families: Evaluation of implementation, context and effectiveness (Family VOICE). Study protocol

Authors

Jonathan Scourfield, Rhiannon Evans, Philip Pallmann, Stavros Petrou, Mike Robling, Kar-Man Au, Delyth Jones-Williams, Fiona Lugg-Widger, Melissa Meindl, Liz Schroeder, Rachael Vaughan, Sophie Wood, David Wilkins

Funding

The study is funded by the National Institute for Health Research Health and Social Care Delivery Research programme. The Children's Social Care Research and Development Centre (CASCADE) Partnership and the Centre for the Development, Evaluation, Complexity and Implementation in Public Health Improvement (DECIPHer) both receive infrastructure funding from Health and Care Research Wales. The Centre for Trials Research receives funding from Health and Care Research Wales and Cancer Research UK.

Competing interests

Jonathan Scourfield is trustee of the Family Rights Group, an organisation that promotes better family participation in the child welfare process, and family group conferences as one way of achieving this. The Family Rights Group are also joint data controllers for the survey of FGC services.

Abstract

BACKGROUND

Family group conferences (FGCs) in child welfare bring immediate and wider family members together to decide on the best way to meet a child's needs. Unlike in professionally-led traditional case conferences, decision-making is shared with family members.

Qualitative and mixed-method research with FGC participants tend to show positive experiences: most participants feel their voices are heard; FGCs facilitate family-driven solutions and closer relationships - within families and with social workers. Although there is existing literature on FGCs, there is a paucity of robust comparative UK evaluations. Comparative studies internationally have focused on a fairly narrow range of outcomes, not recognised the importance of context on FGC effects, and paid little attention to the quality of delivery. Some qualitative studies have considered process and context but there is scant measurement of these.

AIMS AND OBJECTIVES

The aim is to establish how FGCs improve outcomes for families and what factors vary their quality. Given the importance of process and context, an evaluation informed by realist and complex systems approaches is needed.

RESEARCH QUESTIONS (RQs)

- RQ1: What is the extent and nature of FGC provision for child welfare in the UK?

- RQ2: How do families and practitioners see FGCs as impacting on families and what do they think might support or hinder their effectiveness?

- RQ3: How does variation in the quality and implementation of FGCs impact on outcomes and acceptability?

- RQ4: What are the longer terms outcomes of FGCs in terms of service use, e.g. child protection involvement and health care? How do these compare with similar families who have not taken part in FGCs and what are the economic consequences of FGCs?

METHODS

This is a multi-method evaluation, including the following elements:

Work package (WP)1: Identification and modelling of FGCs in the UK; modelling programme theory and development of evaluation design.

- A survey of FGC services in all UK local authorities (n=212) to map service provision.

- Co-production of programme theory and evaluation measures with family members who have experienced an FGC (n=16-24) and practitioners (n=16-24) in each of two sites.

WP2: Prospective single-arm study of FGC variability and outcomes after 12 months.

- A survey of family members (n≈200 families) just before an FCG, one month and 12 months after, to assess process, context and self-reported outcomes; plus a survey of FGC co-ordinators to further assess context.

WP3: Comparison of service use in FGC participants and a comparison group after 2 years using a natural experiment.

- A study of longer-term service-use outcomes and the cost of these, using anonymised and linked routine administrative data on social care and health services. Families who have attended an FGC (n≥300) will be compared with similar families from local authorities with no FGC service (n≥1000).

TIMELINES FOR DELIVERY

The project started in October 2021 and lasts 4 years.

ANTICIPATED IMPACT AND DISSEMINATION

Evaluation questionnaires which have been co-produced with families and practitioners will be available for use in routine FGC practice beyond the life of the project. Findings about FGC process, context and outcomes should be relevant to improving practice. There will be dissemination in accessible formats for practitioners and families, including video, as well as via academic conferences and journals.

The study is registered with Research Registry (ref. 7432)

1. Background and Rationale

1.1 What is the problem being addressed?

A family group conference (FGC) in child welfare brings immediate and wider family members together to decide on the best way to meet the needs of a child who requires support and/or protection. Unlike professionally-led traditional case conferences, FGCs aim to share decision-making with family members. They originated in New Zealand, with a focus on Māori families and in the context of concern about the over-representation of Māori children in state child welfare interventions (Ban, 2005).

There is also considerable current concern in the UK about rising levels of state intervention, especially the year-on-year rising rates of children being 'looked after' by local authorities (Family Rights Group, 2018). There is concern about a professionally-led child protection system which can be confrontational in style (Forrester et al., 2008) and focused on forensic investigation at the expense of support for families (Featherstone, White and Morris, 2014). This critique first emerged in Government-commissioned research in the mid-1990s (Dartington Social Research Unit, 1995) and has not gone away. In this context, FGCs are often seen as an important element of developing a child welfare system that is based on more positive relationships between the state and vulnerable families (see, for example, Mason et al., 2017).

FGCs have the two primary aims of increasing family participation in important decisions about children and reducing more intrusive state intervention. Evidence about the effectiveness of family group decision-making is mixed, however (McGinn et al., 2020; Nurmatov et al., 2020), and variation in quality and context might explain the mixed results.

The study involves a multi-method evaluation. First, the study will involve co-producing evaluation tools with families who have experience of FGCs. Second, a longitudinal study will assess the impact of FGC quality (e.g. achieved participation) and context on short- and medium-term family outcomes. Third, we will compare longer-term subsequent service use with families in similar circumstances who have not taken part in FGCs because these were not offered in their local authority. The theoretical basis of the study is a complex systems approach (HM Treasury, 2020).

1.2 Why is this research important?

FGCs are widely used in the UK. The Family Rights Group website https://www.frg.org.uk/involving-families/family-group-conferences reports that three-quarters of local authorities in England and Wales run or commission an FGC service or are planning to do so. We roughly estimate that around 15,000 children in the UK or more may be involved in FGCs each year. This number is a conservative estimate, based on the referrals expected for the current randomised controlled trial of pre-proceedings FGCs for England (Coram, 2020).

A noted above, FGCs are seen as one way of addressing the over-professionalisation of help for families and as a key element of shifting children's services from excessive statutory intervention to more supportive relationships with families. Mixed evidence about FGC effectiveness may, however, be preventing their wider use. The mixed evidence implies mixed quality of delivery. Better evidence about what constitutes high quality FGCs, in what contexts they thrive, and what outcomes are achieved from high quality practice might lead to more widespread high quality practice, with more benefit to children and families.

The research will establish which aspects of FGCs improve the experience of them for family members and which aspects are linked to better outcomes. A community of practice will be built whereby several existing FGC services will adopt the routine use of a new evaluation questionnaire, which has been co-produced with families and practitioners. The research will also use linked routine administrative data for assessing longer-term social and health care outcomes for family members.

Progress is needed in the theoretical and methodological approaches used to evaluate interventions, in order to understand the variability in outcomes. Realist evaluation (Pawson, 2013), and increasingly complex system perspectives (Hawe, Shiell and Riley, 2009; Moore et al., 2019), maintain there is dynamic interdependence between intervention theories, components and the system in which they operate. Complex systems thinking goes so far as to conceptualise interventions as system disruptions, moving beyond a simplistic focus on the complexity of components and foregrounding the role of context. The present study adopts this complex systems lens in order to understand the link between outcomes, quality of implementation and the systems within which FGCs are delivered.

1.3 How does the existing literature support this study?

Qualitative and mixed-method research with FGC participants suggests that most families who have had an FGC feel that they offer an increased opportunity for their voices to be heard; hence, they can be seen to facilitate family-driven solutions (e.g. Holland et al 2005, Mason et al 2017). FGCs have also been found to help foster closer relationships between family members and improve partnership working between social workers and families (Pennell and Burford 2000, Metze et al 2015). Fewer studies have focused on the experiences of children specifically, but these have also reported positive experiences, as well as some challenges (Bell and Wilson, 2006; Holland and O'Neill, 2006).

It is important, however, to compare families who have attended FGCs with similar families who have not, to see how outcomes and experience compare. Several randomised controlled trials and quasi-experimental studies have been conducted internationally, with 18 included in the recent systematic review by Nurmatov et al. (2020) of all types of participative family meetings. Only two of the FGC studies included in this review took place in the UK (Mason et al 2017; Munro et al 2017) and all the FGC studies had a high or moderate risk of bias. Most focused on reduction in out-of-home care, showing mixed evidence of effectiveness. A similar systematic review for the Campbell Collaboration by McGinn et al (2020) also found mixed evidence for this outcome. Other outcomes are important, given that FGCs are designed to increase family involvement as an important goal in its own right (Morris and Connolly, 2012). Few comparative studies of any type of participative family meeting have considered outcomes such as family satisfaction, family empowerment and engagement with support services (McGinn et al., 2020; Nurmatov et al., 2020) but the evidence on effectiveness for these outcomes is also mixed and of low quality.

Nurmatov et al (2020) conclude that one possible explanation for mixed effectiveness is variation in the quality of delivery. They note that 'it is likely that there is considerable variation in how well shared decision-making meetings, and indeed treatment as usual services, are actually delivered' and they go on to say that 'it is very likely that the passion and skill with which services are provided has a substantial impact on the difference' (Nurmatov et al., 2020: 41). They found that few studies in their review reported on fidelity to the shared decision-making model and none reported on the quality of delivery of usual services. They recommend that future studies include evaluation of the actual service delivery in addition to outcomes.

There is substantial variation in when and how FGCs are used across the UK, with some local authorities using them for early help (Early Intervention Foundation, 2018) and others using them at a very late stage when children are close to coming into care (Coram, 2020). The Leeds Family Valued model, being rolled out by the Department of Education in England as best practice, involves the very widespread use of FGCs (Mason et al., 2017). This includes their use in cases of domestic violence – a fairly novel application (see Sen et al., 2018) and as part of a new model replacing initial child protection conferences.

Although there are strong proponents of FGCs as a fundamental right (see Turner, 2019), they are not without their criticisms. For example, Connolly and Masson (2014) see FGCs as having strayed away from their democratic origins to become instrumental and professionally led. Moyle and Tauri (2016) go further in characterising an approach that was designed in

New Zealand to respect indigenous solutions but has become Eurocentric, formulaic and overly standardised from the perspective of some Māori participants.

Some studies suggest that families who have had FGCs do not differentiate the process itself and their engagement with it from the outcomes they experience (e.g. Frost and Stein 2009). Mitchell (2020) found that family members' personal experience of the FGC process impacts upon their perception of the relevance of support offered, their sense of empowerment and their perspectives on outcomes achieved. The aspect of FGCs often thought to be the most important process issue, namely families' rights to participation being upheld (Morris and Connolly, 2012), can also be conceived of as an outcome (What Works for Children's Social Care, 2020). A few studies have measured family empowerment as an outcome, as in part a proxy for rights being upheld. In the evaluation, we will work with families to establish what for them are the important outcomes of FGCs and assess how well these are met in practice.

Eight randomised controlled trials (RCTs) of family group decision-making, in the USA and the Netherlands, were identified by Nurmatov et al. (2020) and there is a trial currently in progress in England of FGCs at the pre-proceedings stage (Coram, 2020). Compared to these RCTs, the study will consider contextual and implementation factors, such as local authority children's services culture and FGC quality (process), thus taking a complex systems approach, which does not require a randomised design. It will also involve both prospective data collection and longer-term follow-up via routine administrative data.

Despite there being a sizeable literature on FGCs to date, there remain a number of key limitations and evidence gaps. There is a paucity of robust comparative evaluations in the UK context. Comparative studies internationally have focused on a fairly narrow range of outcomes with minimal recognition of the contextual contingency of effects and little attention to the quality of service delivery. Some qualitative studies have considered process, but these have not involved measurement of context and process variables. The research evidence on FGCs demonstrates some of the complexity of delivering the intervention. In light of this, a more nuanced, complex systems approach to evaluation is needed. We propose a study which involves co-production; has a comparative element; and measures context and process as well as outcomes in the short, medium and longer term.

2. Aims and objectives

The study's aim is to establish how FGCs improve outcomes for families and what factors vary their quality.

Research questions:

- RQ1: What is the extent and nature of FGC provision for child welfare in the UK?
- RQ2: How do families and practitioners see FGCs as impacting on families and what do they think might support or hinder their effectiveness?
- RQ3: How does variation in the quality and implementation of FGCs impact on outcomes and acceptability?
- RQ4: What are the longer terms outcomes of FGCs in terms of service use, e.g. child protection involvement and health care? How do these compare with similar families who have not taken part in FGCs and what are the economic consequences of FGCs?

3. Research Plan and Methods

3.1 Design and theoretical/conceptual framework

The study is informed by both realist and complex systems thinking, integrating these into the well-established Medical Research Council (2008) approach for evaluating complex interventions, as demonstrated by Fletcher et al., (2016) and Moore et al. (2019). Theorising intervention process and effects will move beyond the traditional, linear approach to modelling interventions (Kellogg Foundation, 2006) and instead work with complex systems modelling, privileging context more clearly (Rehfuess et al, 2018; Mills et al, 2019). In conceptualising context, we will draw on the framework of Pfadenhauer et al. (2017).

It is a multi-method evaluation, including a survey to map service provision; qualitative co-production of programme theory and evaluation measures; a survey of family members at two time-points to assess process and self-reported outcomes; and a comparative study of longer-term service-use outcomes using linked routine administrative data. The study is registered with Research Registry, with unique reference number researchregistry7432.

Project work packages:

- WP1: Identification and modelling of FGCs in the UK; modelling programme theory and development of evaluation design (RQ1, RQ2)
- WP2: Prospective single-arm study of FGC variability and outcomes after 12 months (RQ3)
- WP3: Comparison of service use in FGC participants and a comparison group after 2 years using a natural experiment (RQ4)

WP1 - Identification of FGCs in the UK; modelling programme theory; and development of evaluation design

WP1 will focus on modelling the programme theory of the FGC intervention, and understanding its local variations within the UK context. This will support the identification of the most appropriate evaluation measures. The work package will address the first two research questions:

- RQ1: What is the extent and nature of FGC provision for child welfare in the UK?
- RQ2: How do families and practitioners see FGCs as impacting on families and what do they think might support or hinder their effectiveness?

Three key activities will be undertaken as part of WP1: identification of FGC services (A); modelling of FGC programme theory (B); and co-production and piloting of evaluation tools (C).

A. Identification of family group conferencing services in the UK

We will conduct a survey of local authorities and their equivalents across the UK to identify how they currently use FGCs, how practice may vary between authorities and what contextual factors may cause such variation.

Sampling:

The sample frame will comprise all local authorities in England, Scotland and Wales, and health and care trusts in Northern Ireland (N=212). As noted above, the Family Rights Group estimate three quarters of local authorities in England and Wales run or commission an FGC service. In Northern Ireland, the Family Group Conference Forum website reports four of the six health and care trusts as having an FGC service (https://www.fgcni.org/). We have not yet been able to find any published summary of the extent of FGC provision in Scotland.

Given that most authorities will have implemented a related service at some point, all will be included in the sample frame. It will also be helpful to know about local authorities who have either never used FGCs and also those who have disinvested in them.

Setting/context:

Local authorities have statutory responsibility for child welfare. They both provide their own services and commission supporting services from the private and third sectors. In Scotland and Wales, there are only unitary authorities, which have child welfare responsibility. In Northern Ireland, the relevant body is health and social care trusts, rather than local authorities. In England, child welfare responsibility sits with 152 authorities, which are either unitary, county councils, metropolitan districts or London boroughs.

Data collection:

An online survey, using Qualtrics software, will be developed in collaboration with the Family Rights Group and in consultation with experienced FGC practitioners (see the section below on modelling FGC programme theory). There will be two routes to reaching FGC services. Firstly, we will send the online survey link to all heads of children's services in the UK and ask them to cascade the survey to relevant staff, such as FGC service managers. Secondly, local independent FGC services will also be identified directly via online searching. There is a list of services on the Family Rights Group website. This list will be supplemented by searching 'family group conference' and each local authority name. Identified FGC services will also be emailed the survey link and asked to take part. Consent will be requested on the first page of the survey.

The online questionnaire will be light-touch, including questions about:

- Whether or not FGCs are used in the local authority/health and social care trust
- The stage of child welfare concern when FGCs are offered
- Criteria for referral
- Which kind of organisation delivers the FGCs
- Is there any theoretical model that informs the approach used?
- How services are functioning in light of Covid-19
- Number of conferences run each year
- What data are currently recorded for evaluation purposes and how valuable this is
- Staff involved including contact details (for WP2)

There will be a mixture of fixed-response quantifiable questions and open questions with free text boxes.

On the basis of previous surveys of local authorities, it is possible that only a modest response rate will be achieved. Baginsky, Ixer and Manthorpe (2020) had responses from 86% of English local authorities for their survey of practice models. However, Corliss et al. (2021) had only 39% response for a survey that was similar in some respects. Even a low response rate would mean a viable study, however, as the main aim is to generate an understanding of different models and different types of implementation, rather than to achieve a highly representative sample. More important would be diversity in context – data from different UK nations and different size local authorities - to capture variation. Additional aims are to make initial contact with practitioners who can form a community of practice for help in delivering WP2 and to identify local authorities where FGCs are not offered, for WP3.

We will work within an adapted overarching ecological framework for research participation (Catherine et al., 2020). We will apply this framework via a separate recruitment and retention protocol for each survey (WP1 and WP2).

For the WP1 survey of services, although we acknowledge the possibility of a modest response rate, we will endeavour to secure as high a response rate as possible. We will dedicate resources and effort to ensure maximum variation to map diversity in practice across different areas. Although the survey should help provide a representative picture of how FGCs are being used across the UK, the more important aim is to indicate key features of use (and non-use) across a maximum variation sample. For both these aims, a high response rate is important.

The recruitment strategy for this survey includes developing a database of services through systematic internet search; using this to initially advertise the study's existence and promote a two-minute video introducing the study; and enlisting the support of groups such as FGC networks and heads of service associations for each of the nations, as well the Family Rights Group which is the only organisation promoting and quality assuring FGCs at a UK national level. Enlisting the support of FGC networks will include study staff promoting the survey and the study more generally to online network meetings where possible. The survey will also be promoted via Twitter.

Data analysis:

Descriptive statistics on survey responses will be produced using Stata software. Free text responses will be thematically synthesised.

B. Modelling of family group conferencing programme theory

This element of the study involves working with key stakeholders – families and practitioners – to co-produce the FGCs programme theory and appropriate outcomes. This will help inform the evaluation design.

We will conduct a series of consultations with stakeholders in order to generate an understanding of the overarching programme theory of how FGCs work in the UK, to construct a complex systems logic model and determine appropriate evaluation measures. Given that there is already a fairly extensive qualitative evidence-base on FGCs (e.g. Holland et al, 2005; Sen et al 2018, Mitchell, 2020), it is preferable for this phase not to include extensive primary data collection but rather comprehensive stakeholder consultation. The consultations will be preceded by a review of the UK qualitative evidence on FGCs by the research team.

Sampling and research participants:

We have identified two FGC sites with contrasting contexts in which to conduct consultation – Camden in London and North Wales. Both areas have FGC services with strong track records of service delivery and ongoing family involvement. These are the Camden Family Group Conference Service and Y Bont, based in Caernarfon, both of which are willing to take part in the research. The two areas are contrasting in several respects. Camden is an inner city, multi-ethnic London borough. North Wales – Y Bont serves Gwynedd, Conwy and Denbighshire local authorities - is a mixture of smaller towns and rural areas, mostly White and containing a high proportion of Welsh speakers, especially in Gwynedd. Both Camden and North Wales have pockets of extreme deprivation. From each of these two contexts we have recruited a peer researcher with experience of having taken part in FGCs as a service user.

Within each of these two FGC services we will purposively sample a group of families who have received a FGC (n=8-12); and a group of local practitioners - both FGC co-ordinators and social workers from children's services (n=8-12). These individuals will be identified and recruited with the help of the local FGC service and the peer researchers, who are locally embedded. Emails and / or direct messages will be sent by the two FGC services to families who have taken part in FGCs, with information about the workshops. People who reply expressing interest will be formally asked their consent to take part. In recruiting participants, careful attention will be given to equality and diversity. Especially relevant diversity issues are likely to be ethnicity, language use and social class. We will ensure the consultation event is pitched appropriately for people without formal education and for ethnic minority participants who speak English as a second language. The Welsh language has equal legal status to English in Wales, so research participation in North Wales will be offered in either language. The workshops in North Wales will include plenary sessions that use a mixture of Welsh and England and separate break-out groups in each language.

Data collection:

In each site, we will host two in-person workshops for adult family members, two online (video conference) workshops for children and young people two online workshops for practitioners. We envisage separate events for family members and practitioners, so that family members feel able to speak freely without social workers present. However, the family events will be run first, so that family members who feel able to join the practitioner consultations can do so. If health protection measures mean in-person meetings are not possible, video conferences will be held.

The first round of consultations will focus on programme theory and the second on evaluation measures. The family events will be jointly run by the peer researchers and other members of the research team. In advance of the first consultation, the research team, including the peer researchers, will draw together data from the survey and the existing research evidence (reviewed qualitative studies and the recent realist review by Stabler et al., 2019) to develop initial candidate logic model(s) that depict the overarching programme theory for FGCs in the UK.

The presentation of the programme theory and logic models will be discussed and agreed upon between the peer researchers and study team in advance of the first consultation events. Part of the discussion will be around accessibility of the language and concepts, as we do not want to put off family members without much formal education or who do not speak English as a first language. The Child and Family Engagement Manager will also help with accessibility. Based on previous experience amongst the study team (e.g. Mannay et al., 2019), we will most likely use creative methods and vignettes. Through the consultations we will explore: 1) Does the initial programme theory fit with the actual experience of being in an FGC? 2) What other things going on (contexts) affect how people experience FGCs (including the effect of Covid-19)? 3) What do good and poorly delivered FGCs look like? 4) What kinds of things do families want to happen as a result of an FGC, that we should be measuring in an evaluation? Note that not all the research questions listed in the plain English summary will be directly used with families in WP1.

We will ask young people who express interest in the workshops about their online connections. If needs be, we will supply a pre-loaded data SIM card to allow them to stay connected. If they are in an area with poor signal and want to take part we will investigate partnering with a local organisation, such as a youth-oriented charity, to provide wifi connection for the young person to participate. We will also explore with the young people that they are able to take part without the control of others. This is part of the rationale for running separate online events rather than young people attending the face-to-face workshop along with adult family members.

In the event that recruited individuals do not want to participate in the group consultation, we will offer the opportunity to participate in an interview. Data will be recorded (audio for in-person meetings and video for online meetings) and transcribed verbatim.

Data analysis:

Data generated through the consultations will be combined with the initial candidate programme theory and logic model(s) to confirm an overarching understanding of FGC practice in the UK, while recognising that not all data will be complementary and a single unified understanding may be unlikely. Such discrepancies will be picked up and explored more fully in the later stages of evaluation (e.g. semi-structured interviews in WP2). Consultation data will be analysed thematically (Braun and Clarke, 2006), using a combination of a priori (e.g. theory context, implementation, outcomes) and in vivo codes. The logic models will be revised so that they are more comprehensive and encompass context and implementation issues. These models will be accompanied by a narrative overview of the programme, which will provide a more detailed account.

Acknowledging that the two case study sites could be considered enthusiasts for FGCs, we will use the WP1 survey of services, which will cover a wider range of FGC practice across the UK, to contextualise our theorising of FGC process in context.

C. Identification and co-production of evaluation measures

We will build on elements A and B above to identify evaluation measures to operationalise our findings for the later stages of the evaluation. Evaluation tools (questionnaires) will be developed and piloted in WP1 to reflect the programme theory, including implementation contexts and outcomes. Following a literature review of existing relevant measures and further consultation with families and practitioners, suitable scale measures will be selected for use in WP2 – this will include a questionnaire for FGC participants at three time points – with existing measures adapted if necessary – and a questionnaire for FGC co-ordinators. The initial selection of potential measures will include peer researchers and options will be discussed at the second round of consultation events with family members and practitioners.

Participation quality is likely to emerge from the first consultation event as a key issue of FGC process. Several scales have been developed for measuring achieved participation, particularly in health care, although few tools that have been designed for completion by patients that provide valid and reliable measures of participation (Philips et al 2016). There are other potentially useful measures that could be used or adapted for this study such as the Family Empowerment Scale (Koren, DeChillo and Freisen, 1992) which has undergone some validation (e.g. Lambert et al., 2020).

The problem of how to deal with multiple perspectives will be explored and resolved with family members and practitioners. Options include separately self-reported measures and a consensus score for the family unit. The acceptability of routine data follow-up for WP3 will be explored with family members, as well as the question of which outcomes discernible from these data sets should be assessed.

The WP1 survey will identify FGC services across the UK. All separate providers will count as individual services, even where there is more than one provider within a single local authority, and the scope will include both services run by local authorities themselves and those commissioned from the third sector (or private sector if relevant). All services will be invited to form a community of practice, to implement the use of the standardised evaluation questionnaires in routine practice in WP2. The questionnaire for FGC participants will be refined in consultation with this community of practice. The questionnaire will then be piloted in one of the FGC services in WP1, using around 20 cognitive interviews to assess validity and likely feasibility. These cognitive interviews will include adult family members, young people and some younger children (in connection with the accessible version of the families' questionnaire).

The questionnaire for FGC co-ordinators - i.e. staff who arrange and facilitate the conferences – will also be developed and piloted in WP1 and distributed as part of WP2. This will characterise the family situation, decisions made and the co-ordinator's own rating of how effective the FGC seemed to be. It will also capture resource use and will be used as a basis for estimating the cost of delivering FGCs. The design of the resource use aspect will be informed by our previous systematic review of economic analyses of FGCs (Nurmatov et al., 2020) and it will be piloted to ascertain its acceptability, comprehension and reliability.

WP2: Prospective single-arm study of FGC variability and outcomes after 12 months

WP2 considers process and also short- and medium-term outcomes of FGCs, using evaluation tools co-produced with families and practitioners in WP1. A questionnaire will be used with FGC participants, before, just after and one year after an FGC. The work package will address the following research question:

- RQ3: How does variation in the quality and implementation of FGCs impact on outcomes and acceptability?

Sampling, settings and research participants

WP2 involves working with a community of practice of FGC practitioners across the UK, identified in the WP1 survey of FGC services, to adopt the families' questionnaire designed in WP1 into routine use. The research team will keep in regular contact with this community. Sites will be purposively sampled for variation in FGC use, with parameters established by the results of the WP1 survey.

Participating FGC services will comprise a prospective single-arm study of the impact of complex systems on family outcomes 12 months after an FGC, using family self-report via a questionnaire and semi-structured interviews with a sub-sample. All family members who have participated in a conference will be invited to take part, with no lower age limit imposed. Those who consent to take part will have contact details recorded at the point of recruitment (baseline) and these will be used for subsequent follow-up. Participants are likely to include extended family members as well as biological and social parents and children / young people.

The evaluation tools will be co-produced with families, so primary outcomes and sample sizes will be decided in WP1. To obtain an estimate of the scale of the study we assume a primary outcome of empowerment and a single composite score across all FGC participants, whilst acknowledging this may not prove to be a preferred measure. We further hypothesise there could be a difference in empowerment according to a characteristic of the FGC – e.g. higher or lower levels of participation or an FGC taking place at a late stage of concern or earlier.

There is not a single agreed empowerment scale used in multiple evaluations. As noted earlier, the Family Empowerment Scale (Koren et al., 1992) is a possible option, but this has not to our knowledge been validated in a UK population. Having consulted results from comparative studies of family group decision-making that have measured empowerment at later time points (Sheets et al, 2009; Dijkstra, Creemers et al., 2018 and Dijkstra, Asscher et al., 2018), it is possible to estimate a sample size for comparing different kinds of FGC. To detect a 0.25 difference between two groups of equal size on a 5-point scale, with a standard deviation of 0.50, 95% confidence and 80% power, a sample size of 128 families will be required (two-sample t-test). For a pre-post comparison to estimate the mean improvement from baseline to 12 months in all FGC participants, N=128 will allow us to detect a minimum effect size of 0.125 with 80% power, or a minimum of 0.144 with 90% power (one-sample t-test). Assuming a 36% attrition rate between the first questionnaire and 12-month follow-up, as was found by Dijkstra, Asscher et al (2018), an initial sample size of 200 FGCs will be needed. The current trial of FGCs at pre-proceedings stage (Coram, 2020) shows an average of 73 conferences a year planned in each of the 24 participating local authorities, so for a minimum of 128 families to participate in the full survey across several different services seems feasible.

Data collection

Survey data will be collected from families at three time points:

- At baseline – i.e. before a conference takes place – data collection will be challenging. Who will actually attend is unpredictable in advance. Once participants have arrived for the FGC but before it starts there are often difficult family dynamics, meaning that completion of a full questionnaire would probably not be feasible. We will assess in WP1 what is possible. Presuming face-to-face conferences are taking place, when participants sign in on paper they can be asked to answer just one or two key questions. In selecting these we would prioritise solely subjective family-reported outcomes where change over time is especially important – an example being empowerment. These brief paper questionnaires can then be posted to the research team. If these are video conferences because of health protection measures, there will be alternative administration methods possible such as email, WhatsApp or text messaging.

- Follow-up at one month after the FGC will capture short-term impacts, in particular, initial family appraisal of the FGC experience and of decisions made;
- Medium-term impacts will be measured at 12 months, to capture enduring consequences of the FGC process for families. Some measures will be repeated from baseline and one-month data collection.

As already established, all questionnaire content will be decided as part of the co-production element of WP1. The questionnaire wording will be as simple as possible, bearing in mind a range of reading abilities, but there will also be a shorter, accessible version for use with young children (e.g. under 11) and people with learning difficulties or low levels of literacy. This will also be co-produced with families and piloted. It will also be checked with a learning disability charity. Administration methods will be determined as part of the co-production of evaluation tools in WP1, but at this point, we can envisage a choice for participants – e.g. online self-completion as the dominant method with researcher-supported data collection over the telephone where requested. We will provide specific training for researchers doing telephone questionnaires, with a script-based approach to maximise reliability.

Soon after each conference, the FGC co-ordinator will fill in a separate short questionnaire (also developed in WP1), to characterise the family situation, decisions made, rating of FGC effectiveness and resource use.

We will draw on the substantial evidence base for maximising response rates and retention (e.g. the systematic reviews by Edwards et al 2002 and Fan and Yan 2010) and consult with families in WP1 about effective approaches. We will offer a financial incentive of £20 per respondent. Our recruitment and retention protocol will include a variety of methods to chase up non-response, including email, text message, and social media. We will ask participants for full contact details and permission to contact them via their preferred method when first recruited, and we will consider requesting secondary contact details (e.g. for a partner), if participants are happy to share these with us. We will also ask for permission to contact participants via the family group conference project or an alternative community group of their choosing if we were to lose contact with them.

Semi-structured telephone interviews will be conducted with family participants, for further exploration of mechanisms and context. We will use a purposive sample of five FGC participants from each of four services, purposively sampled in terms of variation of context – e.g. stage of child welfare concern when FGCs take place and organisational location of services (i.e. run by local authorities or by independent organisations). The interviews will be audio recorded and transcribed.

The intention is for the evaluation tools to become embedded into routine practice, enabling ongoing self-evaluation by services, an essential part of fostering a culture of service improvement. Ongoing support will be provided from research project staff to support the community of practice, with a view to it becoming self-sustaining after the project has finished.

Some important considerations for embedding the research tool into routine practice include co-produced content and data collection method (WP1); liaison with umbrella organisations such as the Family Rights Group, FGC networks for each of the four nations and the British Association of Social Workers FGC practitioners group; collecting no more than families and services can tolerate; a real-time system for monitoring completion; ensuring feedback to service users and FGC staff to show how data are being successfully collected;

preparedness to be flexible, for example in terms of data collection mode preferences; and recruiting local champions for sites to support completion and contribute to problem solving.

We have explored the feasibility of routine use with FGC practitioners and one practical reality to consider is that funders often require some specific evaluation data. In the case of Y Bont, this is different in the three local authorities who commission their FGC service. So any questionnaires for families or practitioners developed in WP1 will need to be appropriately short and focused, as they may well have to be used alongside another type of data collection for local purposes. The logistics of this will be explored with services participating in WP2.

Data analysis

Descriptive summaries of outcomes, by time point, as well as simple pre-post comparisons (mean difference with two-sided 95% confidence interval) will be produced using Stata or R software. Contexts will be assessed per FGC, from a combination of family and practitioner questionnaire data. Additional analyses will involve fitting two-level regression models with random family effects to study the interplay of multiple contextual and implementation factors per follow-up time-point and longitudinally, as well as outcome variability between family members attending the same FGC. Fixed effects will include time-point, contextual / implementation factors and possibly time-by-contextual factor interaction terms. Examples of possible contextual and implementation factors would be family size, stage of the child welfare process (e.g. early help, child in need, edge of care) and local authority culture, as assessed by family and practitioner (possibly a scale measure). Estimates of the effects of contextual factors will be presented with two-sided 95% confidence intervals and p-values.

Data from the WP2 family survey data one month after FGC will be used to assess the reliability and validity of outcome measures by repeating or extending some of the original validation work on key measures, particularly where they have been adapted for use in this study, e.g. Cronbach's alpha, criterion validation and examination of factor structure and ceiling and floor effects. Semi-structured interviews will be coded in N-Vivo software, using thematic analysis.

WP3: Comparison of service use in FGC participants and a comparison group after 2 years using a natural experiment

WP3 is a natural experiment to assess service use outcomes of FGCs and the economic dimension of these. The work package will answer the following research question:

RQ4: What are the longer terms outcomes of FGCs in terms of service use, e.g. child protection involvement and use of health care? How do these compare with similar families who have not taken part in FGCs and what are the economic consequences of FGCs?

Sampling, settings and research participants

Families who have attended FGCs will be identified by local authorities – we will work with several large authorities in England, the Department for Education, NHS Digital and the Office for National Statistics to arrange anonymised data linkage. This part of the study is not linked in any way to the WP2 families. In these participating local authorities we will also collect routine data on families who have been referred to children's services but have not had an FGC, in order to delineate the circumstances in which families do and do not take part in FGCs in these particular authorities.

A comparison group of local authorities which do not offer FGCs will be identified in the WP1 survey. Families from these authorities will be identified from the Children in Need Census. This part of the study uses data from only one UK country – England – because the datasets are different for each nation and given the complexity of data linkage it would not be feasible to attempt this study in more than one nation. We have chosen England as having the largest population.

Eligible families will be those with at least one child under 18 years of age referred to departments of Children's Social Care Services from 1.1.17 to 31.12.19 at participating local authority sites in England. (There is no intention to connect these LAs to those taking part in earlier work packages). No additional exclusion criteria will be applied. At this stage we are not imposing a lower age limit. However, although we are confident that linkage to health records for school-age children is viable, there is some uncertainty about this linkage for pre-school children. We will therefore have to test the feasibility of extending the age profile to include all children from birth, in the early stages of WP3.

To make the two groups of families from different local authorities comparable for analysis, we will use statistical techniques to improve balance across the two groups for a number of known confounders. We will use propensity score weighting (Li et al., 2018) to estimate the average 'causal' effect of having an FGC on the outcomes. The whole of WP3 will be informed by the Medical Research Council's guidance on natural experiments (Craig et al., 2012).

A realistically achievable sample size of 300 families who have had an FGC and 1000 'control' families in local authorities that do not offer FGCs would allow for the detection of a difference of 10.5 percentage points between groups for a binary outcome in a 'worst-case scenario' (i.e. proportions around 0.5) with 90% power and 95% confidence (two-sample test for proportions), but since propensity score weighting tends to shrink the 'effective' sample size the minimum detectable difference will be somewhat larger.

Data collection

Data linkage model: Children referred to children's services in applicable English local authorities (FGC providers and non-providers) during the review period (2017-19) will be identified from the Child in Need dataset within the National Pupil Database (NPD, Department for Education) which can be used to identify all school-aged children. The unique pupil ID can be used to link to other relevant NPD datasets, including Child Looked After. Matching to Hospital Episode Statistics (HES) records will be undertaken using standard linkage strategies for HES and NPD data (e.g. using DOB, gender, postcode) and quality of linkage assessed by estimating rates and distribution of linkage errors. Linkage will be undertaken either by the Department for Education or NHS Digital. Identifiers for FGC families will be supplied from local authorities. Healthcare records for adults will also be matched based on shared household address. The exact model for data linkage will be established in WP1. Anonymised, linked data will be stored in the Office for National Statistics Secure Research Service (ONS- SRS) as a stand-alone dataset and analysed by Cardiff staff via verified remote access.

We will follow up children in these datasets for a period of two years following the date of referral. We have allowed for this work package to run throughout the four years of the project in recognition that permissions for data linkage and the delivery of linked data are very time consuming and subject to delays. We also need to allow for follow-up. The latest follow-up data will be recorded in December 2021 and this data set will be first available to researchers in the Spring of 2023, two years into the project.

The Covid-19 lockdown will of course influence the patterns of service provision from the end of March 2020, but this will be the same for families who have and have not taken part in an FGC and changes in service patterns will be interesting in their own right. Any difference between FGC and non-FGC families in post-lockdown trends can be noted. How Covid-19 and lockdown affect the data will be taken into account when we write a definitive analysis plan. It is not possible to do this until the WP1 co-production phase is complete.

Data analysis

We will estimate a family's propensity for having an FGC with a logistic regression on variables including stage of the child welfare process (referral, child in need, child protection assessment, child protection plan), reason for referral and family demographics. The primary analysis model for a binary outcome variable [to be determined after WP1] will be two-level logistic regression, with random effects for localities and weighting of families according to their propensity for having an FGC. Fixed effects will include intervention (whether a family has had an FGC or not) as well as other factors – to be determined in WP1 - to account for the different ways FGCs are set up in different localities. Effect estimates will be presented with two-sided 95% confidence intervals and p-values.

We will undertake multiple imputation to explore the impact of potential missing data. We will also consider alternative covariate balancing or matching methods, such as entropy balancing (Hainmueller, 2012), Mahalanobis distance matching (Rubin, 1980) and coarsened exact matching (Iacus et al., 2012) if propensity score weighting does not achieve good enough covariate balance and/or as sensitivity analyses. A detailed statistical analysis plan will be finalised prior to any analysis being performed, and published as an addendum to this protocol.

The outcomes for WP3 will be specified after discussion with families and practitioners in WP1, but at this stage we can say they are likely to include child protection involvement (e.g. a composite measure of having a child protection plan and/or being looked after) and aggregate measures of secondary health care service use (e.g. by index child and main carers, for physical and mental health), the components of which will be costed using national tariffs for the purposes of economic analysis. We will consider all out-patient and in-patient episodes, including emergency attendances, using diagnoses and treatment codes. There is no clear hypothesis linking FGC attendance to any single health outcome, but a tentative hypothesis, to be confirmed through modelling work and consultation in WP1, is that an approach emphasising families developing their own problem-solving skills could plausibly be linked to reduced need for health care. We will collaborate with a child and adolescent mental health speciality psychiatrist to advise on codes for mental health conditions which may be especially pertinent to the intervention.

Economic evaluation

An economic evaluation of FGCs will be conducted from a health and social careperspective. Primary micro-costing will be undertaken in WP2 via the practitioner questionnaire to estimate the cost of delivering FGCs across service model delivery types and geographical settings, including the costs of the identification and referral processes, training of accredited providers, delivery of sessions in alternative formats, participant monitoring activities and any follow-up/management. Where applicable the unit costs of resource inputs will be estimated in accordance with the principles of opportunity costs. Broader health care, social care and education resource use over the follow-up period of interest will be extracted from HES and NPD records and costed using national reference cost schedules. The cost-effectiveness of FGCs will be expressed in terms of incremental cost per additional case of child protection or out-of-home care avoided. Bivariate regression of costs and consequences, with multiple imputation of missing data (e.g. loss to follow-up), will be conducted to generate within-study estimates of incremental cost-effectiveness associated with FGCs.

Our analytical strategy will be informed by recent guidance on accounting for selection biases within economic evaluations using individual patient level observational data (Krief et al., 2013). A key methodological challenge will involve generating expressions of cost-effectiveness amenable to broader cost-effectiveness comparisons by decision makers. A range of economic values for avoiding adverse child welfare outcomes will be informed by a review of the revealed and stated preference literature related to children's social care, as well as by WP1. A series of sensitivity analyses will be undertaken to explore the implications of uncertainty on the incremental cost-effectiveness ratios.

4. Project / research timetable

Dates (6m block)	Project year	WP & duration	Main activity	WP & duration	Main activity
Oct 21 - Mar 22	Year 1	WP1 (20m)	– Ethical approval – Survey to identify FGC provision – Programme theory consultation – Evaluation tools consultation – Piloting measures	WP3 (44 months)	– Outcomes discussed with WP1 families – Apply for permissions to link admin data – Admin data sets deposited in safe haven – Data linkage – Admin data analysis and write-up
April 22 – Sept 22					
Oct 22 - Mar 23	Year 2				
April 23 – Sept 23		WP2 (24m)	– Recruit family members from 200 FGCs – Questionnaire follow-up (12m) – Data analysis and write-up		
Oct 23 – Mar 24	Year 3				
April 24 – Sept 24					
Oct 24 – Mar 25	Year 4				
April 25 – Sept 25		Dissemination (4m)			

5. Project management

A Project Steering Committee has been established to bring independent academic and sectoral expertise. The Chair is an experienced social work academic, and the committee includes methodologists, a statistician, the Family Rights Group who quality assure FGCs, and a civil servant working in national child and family policy. It will meet once a year throughout the project. The Project Steering Committee will agree on Stop-Go criteria for the project and assess progress against these. A Families' Research Advisory Group will also run throughout the project, in parallel to the Project Steering Committee. Members will be recruited from existing groups such as the CASCADE families' group and the Family Rights Group's panels, as well as directly from FGC services. This group will also feed into Stop-Go decisions and members will be asked if they want any joint meetings with the Project Steering Committee.

The project will use the Standard Operating Procedures established by the Centre for Trials Research for key processes such as data management and data security.

6. Ethics / Regulatory Approvals

Ethical approval will be sought from the Health Research Authority's Social Care Ethics Committee. This approval will cover the whole of the study, including the use of routine administrative data, but to date only approval for Work Package 1 has been sought – favourable opinion ref. 22/IEC08/0003.

Approval has been given by the Association of Directors of Children's Services Research Group for England and the Northern Ireland Health and Social Care Board. No such national approval exists in Wales or Scotland.

How to achieve genuinely informed consent will be an important consideration for WPs 1 and 2. Anonymity within the study will not be possible for family members participating in WP2, as they will need to input their names and contact details for later survey follow-up. This information will be stored securely on the university server and deleted from the online survey platform. In writing up the results of all work packages, no individuals will be identifiable in any way. If it is important to write up specific aspects of lived experience from WP1, these will be disguised so that no individual can be deductively identified.

To access administrative health care HES records from England without consent, we will apply for Section 251 approval (National Health Service Act 2006, Health Service (Control of Patient Information) Regulations 2002) from the Health Research Authority's Confidentiality Advisory Group (CAG). This temporarily sets aside the common law of confidentiality. As data are still subject to data protection regulations, we will enable opt-out provision via a privacy notice posted on the CASCADE website. The notice will also be available in the five languages other than English or Welsh that more than 200,000 people were reported to speak in England and Wales in the 2011 census – Polish, Punjabi, Urdu, Bengali and Gujarati.

Existing national data opt-outs introduced from May 2018 in England will also be applied by NHS Digital (as data controller for NHS health data). To access social care and health data, we will adapt a model we established for linking from the Department for Education and Department of Health using an accredited data safe haven, building on the approach used by the SAIL Databank to create linked anonymous datasets (Ford et al., 2009; Lugg-Widger et al., 2018). Access to social care data in the National Pupil Database (NPD, which includes the Children in Need and Children looked After data sets) will require approval from the Department for Education and include assessment of compliance with principle 1 of the Data Protection Act (for which the opt-out model described above is included). For both data providers, research in the public interest will be the legal basis for access. NPD and HES data will be linked within either the DfE or NHS Digital data Safe Havens.

Access to social care data sourced directly from local authorities will require Social Care Ethics Committee approval and support from contributing local authorities. Cardiff University will undertake a Data Protection Impact Assessment (DPIA) before any data are accessed or processed. The DPIA is built into the existing Risk Assessment Review used by the Centre for Trials Research (CTR) which will be extended to cover all data collected in the study. Security assurances (e.g. for applications to NHS Digital) will be provided based on the current CTR Data Security and Protection Toolkit return. All individual level linked project data will be stored anonymously at the ONS- SRS and will not be linked to any other data in the ONS – SRS. Accredited researchers at Cardiff University will access data remotely under the terms of the ONS – SRS Security Level arrangements and CTR's own Standard Operating Procedures for remote data access.

7. Impact

The Family VOICE project aims to improve the evidence base informing family group conference delivery in the UK. In terms of impact on practice, the aim is for the WP2 questionnaires to be routinely used by services beyond the lifetime of the study. The research team will work with national networks and key national organisations that promote FGCs, such as the Family Rights Group and Children 1st, in support of this aim. Further workshops will be run to feed back results and discuss implications with family members in the two participating sites and with practitioners. Results of the research will be published in academic journals, as well as formats more accessible for busy practitioners and family members such as short videos. Results will also be presented to academic and practitioner meetings.

Author contributions

Jonathan Scourfield - led protocol writing and study conceptualisation

Rhiannon Evans – contributed to protocol writing and study conceptualisation

Philip Pallmann – contributed to protocol writing and study conceptualisation

Stavros Petrou – contributed to protocol writing and study conceptualisation

Mike Robling – contributed to protocol writing and study conceptualisation

Kar-Man Au – reviewed and edited protocol from a lived experience perspective

Delyth Jones-Williams – reviewed and edited protocol from a lived experience perspective

Fiona Lugg-Widger – reviewed and edited protocol

Melissa Meindl – reviewed and edited protocol

Liz Schroeder – reviewed and edited protocol

Rachael Vaughan – reviewed and edited protocol

Sophie Wood – reviewed and edited protocol

David Wilkins – reviewed and edited protocol, original study conceptualisation

References

Baginsky, M., Ixer, G. and Manthorpe, J. (2020) Practice frameworks in children's services in England: An attempt to steer social work back on course? *Practice*, awaiting issue.

Ban, P. (2005) Aboriginal child placement principle and family group conferences, *Australian Social Work*, 58:4, 384-394.

Bell, M. and Wilson, K. (2006) Children's views of family group conferences, *The British Journal of Social Work*, 36, 4: 671–681.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101.

Catherine, N.L.A., Lever, R., Marcellus, L. et al. (2020) Retaining participants in community-based health research: a case example on standardized planning and reporting. *Trials*, 21, 393.

Connolly, M. and Masson, J. (2014) Private and public voices: Does family group conferencing privilege the voice of children and families in child welfare? *Journal of Social Welfare and Family Law,* 36:4, 403-414,

Coram (2020) *Protocol for a randomised controlled trial of Family Group Conferencing at pre-proceedings stage*. What Works for Children's Social Care.

Corliss, C., Addis, S., El-Banna, A., Maxwell, N., Scourfield, J., Warner, N. and Williams, A. (2021) The views of local authorities in England on how to prevent children being in care. *Child Care in Practice*, awaiting issue.

Cowden, S. and Singh, G. (2007) The 'User': Friend, foe or fetish? A critical exploration of user involvement in health and social care. *Critical Social Policy* 27(1): 5-23.

Craig P, Cooper C, Gunnell D, Haw S., Lawson K, Macintyre S, Ogilvie D, Petticrew M, Reeves B, Sutton M and Thompson S. (2012) Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. *Journal of Epidemiology and Community Health* 66:1182-1186.

Dartington Social Research Unit (1995) *Child Protection: Messages from Research*, London, HMSO.

Dijkstra, S., Asscher, J.J., Dekovic, M., Stams, G.J.J.M. and Creemers, H.E. (2018) A randomized controlled trial on the effectiveness of family group conferencing in child welfare: effectiveness, moderators, and level of FGC completion. *Child Maltreatment*, 24, 2: 137-151.

Dijkstra, S., Creemers, H.E., van Steensel, F.J.A. *et al.* (2018) Cost-effectiveness of Family Group Conferencing in child welfare: a controlled study. *BMC Public Health* **18,** 848.

Early Intervention Foundation (2018) Case Study – Family Group Conferencing Camden https://www.eif.org.uk/resource/family-group-conferencing-camden

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R. and Kwan, I. (2002) Increasing response rates to postal questionnaires: systematic review. *British Medical Journal* 324, 7347, 1183.

Family Rights Group (2018) *Care Crisis Review - Options for change*. London, Nuffield Foundation.

Family Rights Group (2020) Delivering family group conferences during the coronavirus https://www.frg.org.uk/images/200219_FGC-covid-guide.pdf.

Fan, W. and Yan, Z. (2010) Factors affecting response rates of the web survey: A systematic review, *Computers in Human Behavior*, 26, 2, 132-139.

Featherstone, B., White, S. and Morris, K. (2014) *Reimagining Child Protection. Towards Humane Social Work with Families*, Bristol, Policy Press.

Fletcher, A., Jamal, F., Moore, G., Evans, R.E., Murphy, S. and Bonell, S. (2016) Realist complex intervention science: Applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions. *Evaluation* 22(3): 286-303.

Ford DV, Jones KH, Verplancke JP, *et al*. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009;9:157.

Forrester, D., Kershaw, S., Moss, H. and Hughes, L. (2008), Communication skills in child protection: how do social workers talk to parents? *Child and Family Social Work*, 13: 41-51.

Frost, N., & Stein, M. (2009). Editorial: Outcomes of integrated working with children and young people. *Children and Society*, 23, 315–319.

Hainmueller J (2012) Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1), 25-46.

Hawe P, Shiell A, Riley T. (2009) Theorising interventions as events in systems. *American Journal of Community Psychology*. 43(3-4):267-76.

Holland, S., Scourfield, J., O'Neill, S., & Pithouse, A. (2005). Democratising the family and the state? The case of family group conferences in child welfare*. Journal of Social Policy*, 34, 59–77.

Holland, S. and O'Neill, S. (2006) 'We had to be there to make sure it was what we wanted'. Enabling children's participation in family decision-making through the family group conference, *Childhood*, 13 (1): 91-111.

Iacus SM, King G, Porro G (2012) Causal inference without balance checking: coarsened exact matching. *Political Analysis*, 20(1), 1-24.

Koren, P. E., DeChillo, N., & Friesen, B. (1992). Measuring empowerment in families whose children have emotional disabilities: A brief questionnaire. *Rehabilitation Psychology*, 37, 305-32

Kreif N, Grieve R, Sadique MZ. (2013) Statistical methods for cost-effectiveness analyses that use observational data: A critical appraisal tool and review of current practice. *Health Economics*; 22(4):486-500.

Lambert, M.C., Trout, A.L., Huscroft-D'Angelo, J. & Duppong Hurley, K. (2020) Psychometrics of family empowerment scale scores for caregivers of youth transitioning from therapeutic residential care, *Journal of Public Child Welfare*, awaiting issue.

Li, F., Morgan, K.L. & Zaslavsky, A.M. (2018) Balancing Covariates via Propensity Score Weighting, *Journal of the American Statistical Association*, 113:521, 390-400.

Lugg-Widger, F., Cannings-John, R., Angel, L. *et al.* (2018) Assessing the impact of specialist home visiting upon maltreatment in England: a feasibility study of data linkage from a public health trial to routine health and social care data. *Pilot and Feasibility Studies* 4**,** 98. doi:10.1186/s40814-018-0294-4

Mannay, D., Staples, E., Hallett, S., Roberts, L., Rees, A., Evans, R. and Andrews, D. (2019) Enabling talk and reframing messages: Working creatively with care experienced children and young people to recount and re-represent their everyday experiences. *Child Care in Practice* 25 (1): 51-63.

Mason, P. with Ferguson, H., Morris, K., Munton, T. and Sen, R. (2017) *Leeds Family Valued. Evaluation report*. Department for Education.

McGinn, T., Best, P., Wilson, J., Chereni, A., Kamndaya, M. and Shlonsky, A. (2020) Family group decision-making for children at risk of abuse or neglect: A systematic review. *Campbell Systematic Reviews*; 16:e1088.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. and Petticrew, M. (2008) Medical Research Council (2008) Developing and evaluating complex interventions https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/

Metze, R. N., Kwekkeboom, R. H., & Abma, T. A. (2015). The potential of family group conferencing for the resilience and relational autonomy of older adults. *Journal of Aging Studies*, 34, 68–81.

Mills T., Lawton R., Sheard L. (2019) Advancing complexity science in healthcare research: the logic of logic models. *BMC Medical Research Methodology*. 19(1):55.

Mitchell M. (2020) Reimagining child welfare outcomes: Learning from Family Group Conferencing. *Child and Family Social Work*. 25:211–220.

Moore G.F., Evans R.E., Hawkins J., Littlecott, H., Melendez-Torres, G.J., Bonell, C. and Murphy, S. (2019) From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation. *Evaluation* 25(1):23-45.

Morris, K. and Connolly, M. (2012) Family decision making in child welfare: challenges in developing a knowledge base for practice. *Child Abuse Review* 21: 41-52.

Moyle, P. and Tauri, J.M. (2016) Māori, family group conferencing and the mystifications of restorative justice, *Victims and Offenders*, 11:1, 87-106.

Munro, E. R., Meetoo, V., Quy, K. and Simon, A. (2017) *Daybreak Family Group Conferencing: children on the edge of care*. Department for Education.

Nurmatov, UB., Foster, C., Bezeczky, Z., Owen, J., El-Banna, A., Mann, M., Petrou, S., Kemp, A., Scourfield, J., Forrester, D., and Turley, R. (2020) *Impact of shared decision-making family meetings on children's out-of-home care, family empowerment and satisfaction. A systematic review*. What Works for Children's Social Care.

Pawson R. (2013) *The Science of Evaluation: A Realist Manifesto*. London: SAGE.

Pennell J and Burford G. (2000) Family group decision making: Protecting children and women. *Child Welfare* Vol LXXIX, 2: 131-158.

Pfadenhauer, L.M., Gerhardus, A., Mozygemba, K. *et al.* Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework. *Implementation Science* 12**,** 21.

Phillips, N.M., Street, M. and Haesler, E.A. (2016) systematic review of reliable and valid tools for the measurement of patient participation in healthcare. *BMJ Quality & Safety*; 25:110-117.

Rehfuess EA, Booth A, Brereton L, Burns J, Gerhardus A, Mozygemba K, et al. (2018) Towards a taxonomy of logic models in systematic reviews and health technology assessments: A priori, staged, and iterative approaches. *Research Synthesis Methods*. 9(1):13-24.

Rubin DB (1980) Bias reduction using Mahalanobis-metric matching. *Biometrics*, 36(2), 293-298.

Sen, R., Morris, K., Burford, G., Featherstone, B. and Webb, C. (2018) ‘When you're sitting in the room with two people one of whom… has bashed the hell out of the other’: Possibilities and challenges in the use of FGCs and restorative approaches following domestic violence, *Children and Youth Services Review*, 88, 441-449.

Sheets, J., Wittenstrom, K., Fong, R., James, J., Tecci, M., Baumann, D.J., and Rodriguez, C. (2009) Evidence-based practice in family group decision-making for Anglo, African American and Hispanic families, *Children and Youth Services Review*, 31, 11, 1187-1191.

Stabler, L., O’Donnell, C., Forrester, D., Diaz, C., Willis, S, and Brand, S. (2019) *Shared decision-making: What is good practice in delivering meetings? Involving families meaningfully in decision-making to keep children safely at home: A rapid realist review*. What Works for Children’s Social Care.

Turner, A. (2019) Academics voice concerns over What Works Centre’s family group conferences study. *Community Care*, 3rd June.

What Works for Children’s Social Care (2020) Outcomes framework https://whatworks-csc.org.uk/research/outcomes-framework-for-research/