

## Short Reply

# When the 'Golden' Standard Should Be the General Standard: Response to a Commentary on the Use of Randomised Controlled Trials to Examine the Effectiveness of Family Group Conferencing

Hanneke E. Creemers<sup>1,\*</sup>, Knut Sundell<sup>2</sup>, Maja Deković<sup>3</sup>, Sharon Dijkstra<sup>1</sup>, Geert Jan J.M. Stams<sup>1</sup> and Jessica J. Asscher<sup>1</sup>

---

<sup>1</sup>*Forensic Child and Youth Care Sciences, University of Amsterdam, Amsterdam, The Netherlands*

<sup>2</sup>*Medical Management Center, Department of Learning, Informatics, Management & Ethics Karolinska Institutet, Stockholm, Sweden.*

<sup>3</sup>*Child and Adolescent Studies, Utrecht University, Utrecht, The Netherlands*

\*Correspondence to Hanneke Creemers, Forensic Child and Youth Care Sciences, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, The Netherlands. E-mail: H.E.Creemers@UvA.nl

## Abstract

Recently, a critical commentary was published in the *British Journal of Social Work* in which the use of Randomised Controlled Trials (RCTs) to examine the effects of Family Group Conferencing (FGC) was questioned and alternatives were proposed to better examine the effects of FGC. According to the authors, a critical approach towards the evidence provided by RCTs is warranted given the uncontrollability of the social reality of FGC and the generally small sample sizes of FGC studies. We think many of the arguments are based on a misunderstanding of RCTs, while the problems indicated are not solved by using a weaker research design. Because we believe it is a moral imperative to support children and families with evidence-based practice, we feel the need to respond to the misunderstandings and correct them. In addition, we



www.basw.co.uk

© The Author 2016. Published by Oxford University Press on behalf of  
The British Association of Social Workers. All rights reserved.

emphasise the necessity for conducting research that allows unbiased conclusions about the effectiveness of FGC.

**Keywords:** Evidence-based practice, Family Group Conferencing, Randomised Controlled Trials

*Accepted: September 2015*

## Introduction

Recently, a critical commentary was published in the *British Journal of Social Work* in which the use of Randomised Controlled Trials (RCTs) to examine the effects of Family Group Conferencing (FGC) was questioned and alternatives were proposed to better examine the effects of FGC (De Jong *et al.*, 2015). Given that FGC is used in child welfare, where the safety and development of children are threatened, and given the moral responsibility of professionals to choose helpful rather than harmful methods based on the best available evidence, we consider it our responsibility to explain the reasons for conducting RCTs. We want to stress the importance of conducting research that allows for unbiased conclusions about the effectiveness of FGC and to inform the public about the limitations of other research designs that are used to measure effectiveness. RCTs do not come without faults but, to our knowledge, the alternatives are less attractive.

## FGC in child welfare

In child welfare, the model of FGC is increasingly used (Frost *et al.*, 2014). In FGC, an independent coordinator helps the family to gather all parties with an interest in the well-being of a child and its family to make a family group plan that teaches and supports active responsibility of the family and its social network (Merkel-Holguin, 1996). FGC is implemented in child welfare to improve child safety (reduce the risk for child abuse and neglect), prevent or shorten coercive measures (including supervision orders and out-of-home placements) and reduce the need for professional care.

Alarming, whether or not the use of FGC in child welfare achieves these intended results is unknown, as there is, due to a lack of high-quality research, no robust evidence that FGC reduces the serious and complex problems that have resulted in the (enforced) involvement of child welfare (Havnen and Christiansen, 2014). For robust evidence, randomised experimentation is preferred as this allows for unbiased inferences on effectiveness (Farrington, 2003). In randomised experiments,

or RCTs, clients are randomly assigned to either the experimental condition in which they are offered the studied care method or programme, or the control condition in which they receive either care as usual or no care. Random allocation reduces the risk of bias by optimising equal distribution of factors that may influence the results of a method or programme over the experimental and control groups, thereby minimising threats of internal validity. With randomised experimentation, differences between the experimental and control groups in outcomes can be attributed to the studied method or programme, because alternative explanations, including passage of time and pre-existing differences between the groups, are highly implausible (Cook, 2003).

## Controllability of the complex social reality of FGC

In their commentary, de Jong *et al.* (2015, p. 4) argue that ‘the outcome of FGC heavily depends on factors beyond the studied social intervention’, including resistance towards professionals and unexpected events that may have large impact on the results. According to the authors, such factors make the setting in which FGC is organised not as controllable as the settings in which medical and psychological interventions are tested, which would negatively affect the potential of RCTs to determine causality. This is a misconception, since randomised experimentation is developed to account for the influence of known as well as unknown factors on the outcome by balancing the distribution of such factors over the experimental and control groups (Dehue, 1997). Thus, proper randomisation results in equivalent levels of resistance towards professionals and an equal likelihood of high-impact events in families in the experimental and control conditions, provided that such factors are not related to the allocated method. By comparing like with like, alternative explanations can be ruled out. This is not the case in quasi-experimental research favoured by de Jong *et al.* to examine effectiveness, as one can never be sure that comparison groups created by non-random assignment are alike for every relevant factor (Weisburd, 2003).

Aside from factors related to the family context, de Jong *et al.* (2015) argue that intervention characteristics make FGC less suitable for evaluation by RCTs when compared to medical and psychological interventions. While the latter interventions are generally conducted following a prescribed protocol and are relatively identical in nature, FGCs are tailored to the situation and result in a unique care plan. In the last decade, however, numerous calls have been made for greater use of RCTs in everyday practice (as opposed to highly structured situations with well-selected participants with a single problem, ‘lab’ situations or ‘university-based’ trials), to identify effective interventions, including interventions whose delivery varies between participants to address a similar

problem (Allen *et al.*, 2015; Fritz and Cleland, 2003; Zwarenstein *et al.*, 2008). Because external validity is higher in RCTs conducted in everyday practice, their results—when adequately reported (Zwarenstein *et al.*, 2008)—may inform clients, professionals, funders and policy makers on the likely benefits, harms and costs of the intervention when implemented in their own situation.

A final factor compromising the controllability of RCTs on FGC according to de Jong *et al.* (2015) is the attitude of professionals involved with families, usually social workers. Although we agree with the authors that the role and attitude of social workers determine the extent to which the model is used in practice (Sundell, 2000), fidelity problems will also exist in non-randomised trials. To determine the effectiveness of FGC while accounting for attributes of the setting that may influence the results, including method fidelity, researchers should measure characteristics of the implementation.

## The need for sufficiently large sample sizes

The second argument of de Jong *et al.* (2015) that makes them question the value of using RCTs to evaluate the effects of FGC is that large sample sizes are needed, especially given the possible impact of unexpected events. According to the authors, sample sizes in RCTs on FGC are generally too small to produce meaningful statistical results. Although, to date, results from only two RCTs on FGC have been published (Berzin, 2006; Berzin *et al.*, 2008), we acknowledge that the small sample sizes in these studies (thirty-nine and thirty-one families in the experimental group and, respectively, twenty-one and nineteen families in the control group) limit their power to detect small effects of FGC when compared to care as usual, as well as the potential to study moderators of effectiveness. While we agree that power calculations should be provided in research papers (Zwarenstein *et al.*, 2008), we do not share unjustified estimates of sample sizes of over 2,500 families. We estimate that, based on event rates of indicators of child unsafety in child welfare populations (e.g. substantiated reports of abuse, out-of-home placements), sample sizes of sixty-five families per group are sufficient to detect relatively small FGC effects on these outcome measures (e.g. Schulz and Grimes (2005) for sample size calculations in RCTs). Including larger sample sizes than appropriate will not be cost-effective.

## The importance of evidence-based practice

The awareness that decisions in clinical practice should be evidence-based rather than 'authority-based' (Gambrill, 2001) is increasing, but

the resistance to randomised experiments is still great (e.g. Dixon *et al.*, 2014). As a result, there is still a lack of trials on social interventions, including FGC, which impacts on services for hundreds of thousands of children and their families who are receiving care in the absence of supporting evidence. This problem is even more urgent since some theoretically promising interventions have been found to be harmful instead of successful (Dishion *et al.*, 1999; Horigian *et al.*, 2010; Lilienfeld, 2007; Kennedy *et al.*, 2002). As we believe in the importance (for children and families as well as for policy and tax payers) of striving for validated methods/interventions, we argue that randomised experiments should be the standard to examine the effectiveness of FGC. If methods are preferred that lack empirical support, it is imperative to inform the client and, if possible, receive consent.

Although, overall, results from methodologically sound studies on FGC do not verify the presumed superiority of FGCs, neither do they disqualify their use (Havnen and Christiansen, 2014). Furthermore, some of the seemingly discouraging results may be accounted for by services of poor quality, not a fault of the FGC model. A relevant motive for offering families FGCs is that they bring increased transparency to the decision-making process of child welfare. FGCs give extended families and the broader social network of the family the opportunity to make vitally important decisions in matters close to their hearts, and might be a way to level the unequal power relation between child welfare and families, thereby improving collaboration. Nevertheless, children's rights of protection from abuse and neglect should never be compromised. Whether or not the FGC approach constitutes a better alternative from the perspective of children's rights still has to be demonstrated.

## References

- Allen, R. W., Prabjit, K. B. and Lanphear, B. P. (2015) 'Randomized controlled trials in environmental health research: Unethical or underutilized?', *Plos Medicine*, **12**, pp. e1001775.
- Berzin, S. C. (2006) 'Using sibling data to understand the impact of family group decision-making on child welfare outcomes', *Children and Youth Services Review*, **28**(12), pp. 1449–58.
- Berzin, S. C., Cohen, E., Thomas, K. and Dawson, W. C. (2008) 'Does family group decision making affect child welfare outcomes? Findings from a randomized control study', *Child Welfare*, **87**(4), pp. 35–54.
- Cook, T. D. (2003) 'Why have educational evaluators chosen not to do randomized experiments?', *Annals of the American Academy of Political and Social Sciences*, **589**, pp. 114–49.
- De Jong, G., Schout, G. and Abma, T. (2015) 'Examining the effects of family group conferencing with randomised controlled trials: The golden standard?', *British Journal of Social Work*, **45**, pp. 1623–9.

- Dehue, T. (1997) 'Deception, efficiency, and random groups: Psychology and the gradual origination of the random group design', *Isis*, **88**, pp. 653–73.
- Dishion, T. J., McCord, J. and Poulin, F. (1999) 'When interventions harm: Peer group and problem behavior', *American Psychologist*, **9**, pp. 755–64.
- Dixon, J., Biehal, N., Green, J., Sinclair, I., Kay, C. and Parry, E. (2014) 'Trials and tribulations: Challenges and prospects for randomised controlled trials of social work with children', **44**, 1563–81.
- Farrington, D. P. (2003) 'Methodological quality standards for evaluation research', *Annals of the American Academy of Political and Social Sciences*, **587**, pp. 49–69.
- Fritz, J. M. and Cleland, J. (2003) 'Effectiveness versus efficacy: More than a debate over language', *Journal of Orthopaedic & Sports Physical Therapy*, **33**, pp. 163–5.
- Frost, N., Abram, F. and Burgess, H. (2014) 'Family group conferences: Evidence, outcomes and future research', *Child and Family Social Work*, **19**, pp. 501–7.
- Gambrill, E. (2001) 'Social work: An authority-based profession', *Research on Social Work Practice*, **11**, p. 166.
- Havnen, K.J.S. and Christiansen, O. (2014) *Knowledge Review on Family Group Conferencing, Experiences and Outcomes*, Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU West), Bergen (Norway), Uni Research Health.
- Horigian, V. E., Robbins, M. S., Dominguez, R., Ucha, J. and Rosa, C. L. (2010) 'Principles for defining adverse events in behavioral intervention research: Lessons from a family focused adolescent drug abuse trial', *Clinical Trials*, **7**, pp. 58–68.
- Kennedy, S. S., Mercer, J., Mohr, W. and Huffine, C. W. (2002) 'Snake oil, ethics, and the First Amendment: What's a profession to do?', *American Journal of Orthopsychiatry*, **72**, pp. 5–15.
- Lilienfeld, S. O. (2007) 'Psychological treatments that cause harm', *Perspectives on Psychological Science*, **2**, pp. 53–70.
- Merkel-Holguin, L. (1996) 'Putting families back into the child protection partnership: Family group decision making', *Protecting Children*, **12**(3), pp. 4–7.
- Schulz, K. F. and Grimes, D. A. (2005) 'Sample size calculations in randomised trials: Mandatory and mystical', *Lancet*, **365**, pp. 1348–53.
- Sundell, K. (2000) 'Family Group Conferences in Sweden', in Burford, G. and Hudson, J. (eds), *Family Group Conferencing: New Directions in Community-Centered Child and Family Practice*, Hawthorne, NY, Aldine de Gruyter, pp. 198–205.
- Weisburd, D. (2003) 'Ethical practice and evaluation of interventions in crime and justice: The moral imperative for randomized trials', *Evaluation Review*, **27**, pp. 336–54.
- Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D. G., Tunis, S., Haynes, B., Oxman, A. D., Moher, D., CONSORT group and Pragmatic Trials in Healthcare (Practihc) Group (2008) 'Improving the reporting of pragmatic trials: An extension of the CONSORT statement', *BMJ*, **337**, p. a2390.