Contemp Fam Ther (2015) 37:1–12
DOI 10.1007/s10591-014-9314-6

ORIGINAL PAPER

Couple and Family Therapy Outcome Research in the Previous Decade: What Does the Evidence Tell Us?

Peter Stratton · Emma Silver · Natasha Nascimento · Liz McDonnell · Gwen Powell · Ewa Nowotny

Published online: 2 October 2014

Abstract Meta-analyses of randomized control trials include only a small proportion of the published outcome research of Couple and Family Therapy. This paper surveys the ranges of published research through a systematic review of the outcome studies of family, couple, and systemic therapies published in English language peer reviewed journals in years 2000 through 2009. After application of criteria of relevance to Couple and Family Therapy and systemic practice, 225 studies were identified, summarized, and coded under 14 broad headings giving 125 potential classifications for each article. Analyses of these codings found consistent conclusions of effectiveness; differential availability and quality of research for different conditions; and quite frequent absence of important methodological information. The findings are interpreted as showing that this body of recent research supports claims of effectiveness. Although the journals included many of good quality there are substantial areas of weakness in reporting. It is concluded that there are significant influences on the body of published research that arise both from funding policies and journal practices as well as perhaps author bias. The consequences are to reduce the value of research to practitioners, to favour randomized control trials with positive evidence of the effectiveness of therapy, and to exclude publication of negative findings.

Keywords Family therapy · Couple therapy · Systemic practice · Outcome research · Evidence · Review

P. Stratton (✉) · G. Powell
Institute of Health Sciences, University of Leeds, Leeds, UK
e-mail: p.m.stratton@ntlworld.com

G. Powell
e-mail: g.powell@leeds.ac.uk

E. Silver
The Brandon Centre, London, UK
e-mail: EmmaSilver@brandoncentre.org.uk

N. Nascimento
Marlborough Family Service, London, UK
e-mail: nascimentonatasha@yahoo.com

L. McDonnell
Department of Sociology, University of Sussex, Falmer, UK
e-mail: thomday@hotmail.com

E. Nowotny
Camden and Islington NHS Foundation Trust, London, UK
e-mail: u1331803@uel.ac.uk

Introduction

There have been consistent claims of the effectiveness of Couples and Family Therapy (CFT) interventions from meta-analyses and systematic reviews (Hazelrigg et al. 1987; Retzlaff et al. 2013; Shadish et al. 1993; von Sydow et al. 2010, 2013) as well as from broader surveys (Carr 2014a, b, Sexton et al. 2013; Sprenkle 2012; Stratton 2011). However, while some manualised therapies lend themselves to rigorous randomized control trials of their application to clearly diagnosed conditions, there are other forms of therapy for which the evidence base is much more varied. Couples and family therapy is one example, having a substantial evidence base but with only a small proportion being provided by studies suitable for inclusion in meta-analyses. The steady accumulation of high quality evidence has succeeded in having CFT included in some international treatment guidelines, for example the American Academy of Child and Adolescent Psychiatry (AACAP) practice parameters for most child and adolescent problems and the UK's National Institute for Health and

Clinical Excellence (NICE) treatment guidelines for eating disorders, schizophrenia, and substance use. However, many areas of practice are supported by research that does not meet the specific requirements of these bodies but there is no evidence that they are less effective.

This review arose from a particular concern that many CFT practitioners do not use the research base to improve their practice, raising the question of whether the forms and topics of research correspond well to the practice and concerns of therapists. It was designed to indicate the full range of recent outcome research in order to discover the areas in which research has been directed, the types and quality of research methodologies employed, and the quality of reporting. In order to conduct the survey, criteria were established for which articles to include. A wide definition of CFT was adopted to ensure representation of the outcome research that would be of value to family therapy practitioners.

'Couples and Family Therapy' incorporates a range of therapeutic interventions (e.g. marital/couple, family, multi-family groups, systemic therapies) with exceptionally wide application (e.g. adult, child and adolescent mental health, behavior problems, drug and alcohol, schools, physical illness, relationship difficulties). Most CFT models share the assumption that the presenting difficulties are inextricably tied to the relational patterns of behavior of the systems in which they exist, usually the family (Dallos and Draper 2010). Therefore, CFT interventions focus on the interpersonal system to promote systemic change and change in the individual presenting difficulties.

Although early reviews of CFT research found positive effects and indicated effectiveness for the main models of CFT available at that time, they also reported weak methodology of the research trials. Hazelrigg et al. (1987) found only 20 of 281 studies that met their criteria for good methodology. However, when Shadish et al. (1993) included only trials with random assignment and distressed participants, it left them with just 71 CFT studies over a 25 year period (1963–1988). More recently, Shadish and Baldwin's (2003) meta-analysis of 20 meta-analyses of CFT indicated that the average treated family fared better after therapy and at follow-up than more than 71 % of families in comparison groups. They concluded that 'Marriage and Family Therapy is now an empirically supported therapy in the plain English sense of the phrase—it clearly works, both in general and for a variety of specific problems.' (p. 567). Sexton and Alexander's (2002) evaluation of family based interventions reported that both the qualitative and meta-analytic reviews provided "substantial support for the general efficacy of family-based interventions". Carr (2014a, b) included reviews of studies that used a range of methodologies, including some case studies, and concluded that the evidence supports the effectiveness of CFT interventions for a wide range of child-focused and adult-focused problems, either alone or in conjunction with other treatments.

Despite the widespread use of CFT, there are few reviews of outcome studies covering the breadth of these interventions across the range of their contexts and applications. At a time when the application and types of CFT are expanding (Carr 2012; Nichols 2012), and in a context of developing evidence based practice, this review explores not only where the outcome research effort has been directed over a recent ten-year period, appraises and synthesizes diverse research evidence, and discusses the findings and implications for clinical practice and research, but also, whether the recent research supports the indications of effectiveness from earlier meta-analyses and broad surveys.

A meta-analysis would not have been appropriate because in the chosen period there is not a large enough group of substantial randomized controlled trials (RCTs) clustered in relation to a single condition and form of therapy. Instead there is a large variability of the methodology of the outcome studies identified. This range in methodology reflects a large number of studies using research methods that are more suited to evaluating outcomes of CFT and in particular the range of circumstances to which CFT is applied in clinical practice. The methodology of this review falls between a broad systematic review, and a narrative review, as generally "Systematic reviews are generated to answer specific, often narrow, clinical questions in depth." Whereas "Summaries of research that lack explicit descriptions of systematic methods are often called narrative reviews. Most narrative review articles deal with a broad range of issues related to a given topic rather than addressing a particular issue in depth." (Crowther and Cook 2007, p. 493). Perhaps the most informative descriptor is that adopted by von Sydow (1999) and von Sydow et al. (2010) of a 'meta-content analysis'.

The review is intentionally pluralistic, incorporating different study designs that investigated a wide range of presenting problems in adults and children, conducted in various clinical contexts. The research included in this review is unusually varied and not tied to specific diagnoses or methodologies, as one of the main aims of the study was to gather information to map out the forms that outcome research in CFT has taken over the previous decade. Until now, most reviews of CFT outcome research have been organized according to particular diagnostic groups, for example, adults with mood disorders (Carr 2014a, b; von Sydow et al. 2010), but in order to capture the wide range of applications of CFT, this review started from looking more broadly at what sorts of CFT research

have been conducted with whom, where, how, and its reported efficacy.

Definition of Family Therapy

A broad definition of Couples and Family Therapy has been used, which covers different types of couple and family therapies and other family-based interventions where family members or the families' wider networks were engaged in the intervention. All of the studies were examined and only those with some aspects of family therapy, or interventions based in systemic theory, with measured outcomes of ameliorating or preventing difficulties were included in the review.

By including a broad range of therapeutic interventions with families, couples, groups and schools, we have reported on therapies from a range of theoretical approaches, which differs from previous recent reviews of CFT outcome studies (e.g. von Sydow et al. 2013). Our broader approach is in line with Lebow and Gurman's (1995) review. It is consistent with the development of theory and practice of CFT, which in common with psychotherapy generally, is expanding to include concepts and techniques from other theoretical models (Norcross and Beutler 2011). However, in order to explore this concept further, the researchers have gone on to categorize each of the studies according to the strength of systemic influence on the therapy in each study, and what form of CFT was being conducted (Silver et al. in preparation).

Objectives

This review explores the characteristics of outcome studies on therapeutic interventions with families and couples published over a 10-year period, 2000 through 2009. The review includes a broad range of studies and reports on the following aspects of family CFT outcome studies: types of study design, range of populations and presenting problems, where the research effort has been conducted, types of therapy/interventions, and range of outcome measures used. In collating this information, this review also examines the quality of evidence in family therapies, both in relation to the claims of effectiveness, and the quality of the evidence provided.

Methods

Literature Search and Study Selection Criteria

A search of selected electronic databases (2000–2009): MEDLINE, PsycINFO, AMED, British Nursing Index and Archive, CINAHL, and EMBASE, was undertaken using the search terms: "Family Therapy", "Couples Therapy", "Systemic" and "Marital Therapy". Carrying out this exercise prior to developing the final search strategy enabled us to compile a list of further keywords associated with CFT that could be included in the final search strategy. The final search was undertaken in a total of 15 databases refined by use of search terms appropriate to each database. A listing of databases and examples of search terms are provided in the Appendix.

Web of Science (2000–2009) and the Cochrane Library (2000–2009) were also searched. The existing search was cross-checked against the reference lists of relevant Cochrane and other reviews and selected referenced articles to identify any studies that we had not previously extracted. A hand search of key journals was carried out. In addition selected experts in the field were sent preliminary listings to check against their specialist knowledge of the literature.

Selection of the articles generated by the searches was conducted in three stages; first by relevance of the title, then by relevance of the abstract, and finally, those that remained potentially relevant were obtained in full text for a more detailed appraisal. All decisions for excluding papers were logged.

Inclusion Criteria

Interventions were with a family or couple, or systemic oriented interventions with schools, groups or individuals; Studies reporting outcome evaluations of CFT interventions; and Articles published in peer reviewed journals in the English Language between January 2000 and December 2009.

Exclusion Criteria

Single case studies; Review articles; No outcomes reported.

Each database searched generated substantial numbers of articles that were clearly irrelevant. After eliminating these, 386 articles remained that apparently fitted the criteria, and these were hand searched independently by two members of the research team to ensure that none were missed, or others included that should be excluded. A further 166 were eliminated. These were mostly review articles that did not report original research, or articles that for other reasons did not report primary data pertaining to outcomes on the clinical effectiveness of CFT interventions/services. The search identified 225 outcome studies that fully met the inclusion criteria and these are included in the review and coded as described below.

Coding and Analysis

Coding Process and Inter-Rater Reliability

To facilitate data collection, we developed a coding protocol for analyzing articles that fitted the inclusion criteria.[1] This was developed in consultation with previous reviews of psychotherapy research (Froyd et al. 1996; Sanderson et al. 2009) to guide the inclusion of the categories. In order to maximize rater reliability, definitions of coding categories were compiled and two researchers independently rated several articles that were cross-checked, and where there were discrepancies, the categories were discussed and re-defined. This process was repeated until a satisfactory level of agreement between the raters was reached. Two members of the research team (ES and NN) each rated 136 articles on all variables. This allowed for 15 % of the articles to be co-rated. Inter-rater reliability was calculated for all coding categories as percentage agreement between two coders as follows: authors' conclusions—96 %; types of outcome measures—95 %; type of therapy—94 %; quality of randomization—91 %; criteria for diagnosis—85 %; ethnicity—81 %; setting—81 %; who was in the session—75 %, and who was measured—75 %. The lower areas of agreement between coders was primarily a reflection of the descriptions in the studies, as it was often difficult to ascertain ethnicity of clients, what the therapy actually entailed in terms of which family members were in the session, where the sessions were conducted, and who was measured. Inability to agree on a classification is included in this form as the areas in which it is not possible to make a reliable judgement from what was reported in the article, which is one of the prime concerns of this study.

Study Type and Quality of Evidence

All the included articles were coded according to study design, presenting problems, sample characteristics, types of interventions, outcome measures, and on criteria relating to quality of evidence as detailed below.

Research Design

Each study was given a dichotomous coding to indicate which design(s) it fitted of the following headings: Randomized Control Trial (RCT—patients/families are randomly allocated to different groups offering different interventions, enabling a direct comparison of outcomes); follow-up of a RCT, quasi-experimental (non-random assignment of participants into comparison groups); pre-post (measure of change over time in one group); qualitative (using qualitative outcome measures only); Case Series (measuring outcomes of more than 3 cases); and Retrospective Study (e.g. looking back at whether therapy was effective from medical records). In addition to being coded as one of the above types of study, some articles were also coded as a "Preventative Study", for example, exploring the impact of a family intervention where a carer/parent was the 'identified patient', measuring the outcome in the children, or vice versa. Although 27 studies were initially identified by their authors as "Pilot Studies", these were re-classified according to the methodology into one of the above categories. Most of those described by authors as 'pilot studies' had more than 25 participants (mean = 49.9), and may have been described in their title as a 'pilot' to meet the requirements for funding and/or publication.

Presenting Problem The studies were allocated to 16 categories according to age group (children up to 18 years, or adults), presenting problems, and diagnoses (based on DSM IV categorization). These included behavioral problems, mental and physical health diagnostic categories, and relationship problems. Generally, studies were categorized according to the presentation of the identified client, rather than the area of intervention. So, for example, in some studies, although the intervention was with the parents, it was the child who was referred with the presenting problem. The categories are provided in Table 2 (below).

Sample Characteristics We recorded: Number of participants in the study (and number in each comparison group); mean age of participants in each treatment group; gender; and ethnicity.

Types of Intervention The methods of intervention used in treatment groups were coded with respect to the named model of therapy reported in the article (such as Behavioral Couples Therapy, Multi-Systemic Family Therapy, etc.). However, for many studies the therapists used a range of methods of intervention, taken from more than one systemic model, as is common practice when working with families (Rivett 2008). The researchers read the papers and often followed up references to attempt to categorize the type of CFT described in the articles. They referred to specific CFT texts, including Carr (2006), Vetere and Dallos (2003), Pote et al. (2003), and used their knowledge and experience of systemic therapies, as well as the research experience of the first three authors in developing the UK Government framework of systemic competences (Stratton et al. 2011a) to label the type of family intervention described. Information on the treatment setting (e.g. inpatient, community, etc.) and who was involved in the session (e.g. child, family, parents, etc.) was also recorded.

[1] The coding sheet is available upon request from the first author.

The comparison intervention was coded according to whether it was another type of family/couple intervention, individual therapy, treatment as usual, no treatment, combined treatments, medication, group therapy, or psychoeducation.

Outcome Measures All studies were coded for who and what was measured as outcomes, and these were categorized as: Objective Measures: e.g. weight, urine sample, recidivism, re-admission to hospital; Self-Report: adult, child/adolescent, parent/carer's measure, client satisfaction; Behavior reported by others: child's/adolescent's/partner's behavior reported by other; Relational Measures: dyadic couple relationship, dyadic parent/carer–child relationship, family systems measures; Other Measures: therapist measure (e.g. adherence), cost of treatment measure; No clearly defined outcome measure.

Quality of Methodology Each study was rated on seven aspects relevant to the quality of the methodology and credibility of the findings. Aspects were only used when reliably reported in a substantial proportion of the papers. So, for example, the level of training of the therapist which has been shown to be a crucial factor in effectiveness by Crane and Payne (2011) was not sufficiently well reported to be included.

- the level of reported randomization (fully reported; claimed with some detail; acknowledged; or not randomized);
- the criteria used for diagnosis (i.e. established criteria for diagnostic process by clinician using standardized measure; self-report using standardized measure; informal/unclear);
- total number of participants;
- whether there was one or more comparison groups;
- use of standardized outcome measures;
- length of follow-up;
- the quality of the outcome measures used (which were rated hierarchically with Objective Measures at the top, then Self-Report, Relational Measures, and lastly 'no standardized measure').

These 7 aspects were combined arithmetically to create an overall measure of quality that could be used in the analyses for comparison with other aspects of the studies.

Results

A descriptive summary of each of the 225 articles and a full listing of the references included in this review is available in (Stratton et al. 2011b) or from the first author.

Table 1 Frequency with which therapy was rated as effective for the most commonly researched conditions

Condition	Proportion rated effective	Percent
Adult substance abuse	28/29	97
Child behavioral problems	25/29	86
Child eating disorder	13/16	81
Adult schizophrenia and psychosis	18/23	78
Adult—other psychiatric	10/13	77
Adult mood (depression)	7/20	35
Adolescent substance misuse	6/20	30
Child anxiety and mood	6/23	26

Do the Findings Support the Indications of Effectiveness in Existing Reviews?

As already discussed, the variety and quality of methodologies and of conditions treated made it impractical to carry out an objective evaluation comparing the effectiveness of the therapies reported. It is however of interest to explore the claims that authors made about the outcomes. These claims can be considered to have some validity as they had been subjected to scrutiny and accepted by the refereeing processes of the journals.

No study reported that the clients deteriorated during therapy, 8 % did not make a clear claim; 16 % reported no significant difference with the comparison treatment; and the majority (75 %) claimed that the therapy was effective. Only 18 % claimed clear superiority over the comparison treatment. Some studies made the comparison with another type of family intervention so they might claim it was effective as an intervention, but only the same as the other type of intervention.

Claims for effectiveness varied with the presenting problem. Considering the more frequently researched problems, the strongest claims of family therapy being effective and/or better than the comparison treatment were as shown in Table 1.

It appears that the weakest claims of effectiveness were for affective disorders, in both child and adult studies. Adolescent substance misuse, which could be considered to have at least a component of affect, also has weak claims. The methodologies rated strongest (see 'quality of methodology' below) were more likely to have weaker author claims. Effectiveness was rated the same as comparison in 20 instances, compared to an expected value of 11.4. Either the more rigorous researchers were more cautious in their interpretations or perhaps the more thorough research gave weaker findings.

Table 2 Studies grouped in broad adult or child diagnostic categories

Presenting condition	Number of published studies	As percentage of all studies reviewed
Adult mood disorders (depression, bipolar)	20	9
Adult substance misuse	29	13
Adult psychosis (schizophrenia)	23	10.2
Adult eating disorders (anorexia, bulimia)	3	1.3
Adult other mental health (post traumatic stress disorder, child sexual abuse, dementia)	14	6
Adult physical health (cancer, pain, HIV)	7	3
Adult relationship problems	13	6
Child/adolescent mood disorders (depression, bipolar)	9	4
Child/adolescent substance misuse	20	9
Child/adolescent anxiety disorders (separation anxiety, obsessive compulsive disorder, post traumatic stress disorder)	14	6
Child/adolescent eating disorders	16	7
Child/adolescent physical illness (cancer, obesity, epilepsy, HIV, asthma, diabetes)	20	9
Child/adolescent learning disability	2	0.9
Child/adolescent behavioural problems	29	13
Child/adolescent other or mixed presentations	4	1.8
Cost effectiveness studies	2	0.9
Total	225	100

The strength of claims of effectiveness differed according to the comparison. When the comparison was with no therapy 78 % claimed effective and/or better than comparison. Treatment as Usual was similar (80 %), whereas comparisons with individual therapy (70 %) and a different form of CFT (68 %) were weaker.

Does Research Reflect the Range of Applications of FT in Practice?

Clinical Areas Studied

Table 2 presents the studies categorized according to the nearest diagnostic category of the identified or referred client (RC), i.e. the family member who brought the family to the service. People aged over 18 years are classified as 'adult'.

The mean age for adults (over 18 with range up to 77 years) was 37.43 years (SD = 10.22, n = 84). The mean age for children (up to 18 with range from 2 years) was 12.54 (SD = 3.65, n = 93). The gender split in the adult samples was equal but in child/adolescents was 56 % male arising from the focus on adolescent delinquency and substance misuse.

Research was concentrated around ages 15 and mid 30 s. It seems likely that this is a narrower focus than general clinical practice and other areas of common clinical practice were under-represented. In only 7 (4 %) studies was the IP over 50 years old, while only 3 studies reported IPs between 16 and 21 years. Three studies concerned adults over 65, two with dementia, one with cancer. Two studies concerned children with learning disabilities and no studies were found with adults with learning disabilities. Overall, with adults there was more research on substance misuse and psychosis than mood disorders (depression and anxiety). Children research on behavior problems and substance misuse far outweighed mood disorders.

Contexts of Application

Only a quarter of the studies were conducted in specialist settings (e.g. research centers or centers of excellence), with 22 % of studies being conducted in the community and 13 % in the home. Other contexts included schools and phone surveys. This is important for clinicians who often have concerns that large clinical trials conducted in expert research contexts or specialized services are not applicable to their clinical work. This review indicates a wide range of contexts for the studies, so the concern about efficacy but not effectiveness may not apply strongly in the field of CFT research.

This review only included studies published in English language journals, so these results are not representative of the full international research product. They may, however, give some indication of the strength of CFT in different parts of the world. North America dominates with 63 %, 133 of these 142 articles originating in the USA. Fifteen of Europe's 50 articles originated in the UK and 16 elsewhere in Northern Europe. Of the 10 Asia papers, half were from China. We found no research from France published in English, whereas Germany, Italy, Spain, Belgium, Switzerland and Ireland all contributed. These results give some indication of the availability of CFT around the world but must partly reflect the availability of research funding in different countries as well as the motivation and capability to write for English language journals.

Nature of Research

Types of Study Design

Considering the methodologies of the 225 studies reviewed 73 % were Randomized Control Trials (RCTs),[2] of which 14 were follow-ups of earlier RCTs; 15 % were non-controlled outcome studies; 4 % were quasi-experimental and 3 % were preventative. Very small numbers used other methodology; only 5 (2 %) using a Qualitative Analysis, 4 were Retrospective Studies, and 4 were Case Series. The mean number of participants in the studies was 106 (number of families/couples rather than individuals within families), with 90 % of studies comparing at least two intervention groups, and 22 % comparing three or more groups. Twenty seven percent of the studies compared the CFT intervention with a treatment as usual (TAU) group; a similar number of studies (24 %) compared the family intervention with another type of CFT; 22 % compared family intervention with an individual therapy (most commonly Cognitive Behavioral Therapy); 5 % with groups (either psychotherapeutic groups or psycho-education), and only 3 % of the studies compared CFT with medication only. Comparisons with medication tended to use combined treatments (medication and a form of psychotherapy) where medication is likely to be part of usual treatment (with adults with severe depression, diagnoses of bipolar disorder, and schizophrenia).

Outcome Measure

Consistent with the findings of Sanderson et al. (2009), there was great diversity in the outcome measures employed and most studies used more than one type of outcome measure. In consequence, the data reported here are not cumulative. 41 % of studies used an objective measure (e.g. weight, urine sample, recidivism, re-admission to hospital), which reflects the large proportion of studies on substance misuse (22 %), delinquency and conduct disorder (13 %), and eating disorders (8 %). A substantial proportion of the studies (79 %) included self-report measures (usually questionnaires or structured interviews with the client), and one quarter used other people's report (e.g. parent's report on child's behavior), which again was in the form of structured interviews or questionnaires. Although all of these studies were with families or couples and many claimed to be conducting systemically informed interventions, only 58 % of the studies gathered measures from more than one person in the family; less than half of them (46 %) included relational measures, and only 25 % used a family system measure. The most commonly used family system measure was in the form of child/adolescent's self-report (33 %).

The majority (91 %) of published studies reported having used a standardized measure of some type of outcome, and 64 % reported use of established criteria for diagnosis through clinician administration of an objective measure (e.g. a diagnostic interview that had been shown to have good reliability and validity to meet DSM diagnostic criteria). Nineteen percent included using established measures through self report of patients. Sometimes CFT, like other psychotherapies, has argued against the appropriateness of objective and standardized measures. The high level found in this survey indicates that this self-imposed limitation is not prevalent among those who research CFT, at least, among those who get published.

Quality of Research

We were interested to investigate the quality of the methodologies used in the research surveyed. In this section we consider the quality of the research judged from the published accounts. The subsequent section explores the quality of the reporting in the journal articles.

One indication is the quality of the journals in which the research was published of which the most direct indicator is the impact factor (IF). This figure was available for the year 2009 for 205 of the articles. The median IF was a respectable 2.81 and the mean 3.44 (SD 2.86) with a range from 0.32 to 14. This is an acceptable average level of impact and does not suggest that outcome articles in family therapy are disproportionately published in low impact journals. In fact, many of the lower IFs (<1) were of specialist journals such as The American Journal of Family Therapy; Journal of Family Therapy and Journal of Marital And Family Therapy which are appropriate vehicles for this material.

A direct measure of the quality of the research methods was obtained by computing a composite measure of the seven aspects of research quality that could be reliably estimated in most cases. We did not want to assume that quality equates with conformity to RCT specifications, as one purpose of this survey has been to determine the range of methodologies employed in investigating the outcomes of CFT and to comment on their quality. The seven criteria are listed above in 2.2.2.6 Quality of Methodology.

Each of the seven criteria was allocated one point and these were summed to give the total score, to a maximum of seven. Where several levels of a criterion were classified, the point was partitioned. For example, randomization was scored 0 if there was no indication of randomization; 0.3 if there was a claim but no detail; 0.6 if information about the randomization was reported but either with

[2] All frequencies are based on those studies for which this aspect could reliably be judged (generally a very high proportion).

limited information or the randomization did not meet full standard criteria, and 1 for full randomization comprehensively reported. Where a study could score on more than one level of the criterion (e.g. by using several different outcome measures) the strongest aspect was scored.

In practice the criteria can all be seen as relevant to the quality of an RCT but the process differs from that of a meta-analysis. Instead of using any of them as exclusion criteria, for example with high requirements for randomization before a study can be included, but less stringency for other indicators, we have given equal weight to each of the measures. A study could therefore have a high quality score even if there was no randomization, if all other criteria were high.

The composite quality measure for each study correlated significantly with the IF of its journal ($r = 0.365$, $N = 201$, $p < .001$). Twenty three articles (10.5 %) were rated at the top level on the measure and these had a higher IF (Mean = 6.15) than the remaining articles (mean = 3.13). So the articles with higher composite quality scores tended to be published in higher rated journals. However, quality of randomization had a higher correlation with journal IF than the other six quality criteria combined. There appears to be a clear tendency for journals with higher status to give far more weight to standards of randomization than to other aspects of quality.

One hundred and fifty of the studies were designated as RCTs by their authors. These covered the full range of quality scores from 0.29 to 6.6 with 61 scoring less than 4.5. In only 62 of the 150 was the quality of randomization rated as 'fully reported', or 'claimed with some detail'. So some designated RCTs were rated as lower quality than other designs and some provided minimal information about the randomization. In order to assess the quality of research independently of the quality of randomization, this variable was removed from the calculation of quality. With this version of the quality measure, the design with the highest mean quality score was RCT follow-up (quality average = 4.41) followed by RCTs (4.10). In order, the other designs achieved quality measures of: Preventative studies 3.83, quasi-experimental 3.56 and non-controlled 2.77. The other three designs were too infrequent to rate separately but had a combined average quality of 2.1.

Standards of Reporting

There was a range in the quality of reporting. Those that clearly described the method, participants, results, and conclusions were easy to code and we could include all the available data in our analysis. However, for some studies it was difficult to find the information, and for others important information about the research was not available in the published article.

For example, 40 % (91) of the papers included in this review did not report on the ethnicity of the participants, 20 % did not report on the mean age of the index patients, and 16 % did not include information about the gender of the participants. Given that sensitivity to gender, culture, and ethnicity is particularly important in family research because they provide a context for understanding the relationships and interactions that are of interest (Sexton and Alexander 2002), we were surprised by this lack of basic information in so many studies.

In general, for many studies it was difficult to label the type of CFT /intervention, or to know where the therapy took place, and who (which family members) were included in the sessions. As this was often not clearly stated by the authors, we elicited some of these data from other information in the article, for example the information about types of therapies described in the introduction and references to named therapies or theoretical models that the intervention was based upon. However, some information was not available in any form. Fourteen percent (31) of the studies did not describe where the therapy took place (i.e. home, hospital, community clinic, etc), 6 % (13) of the studies did not describe which model of therapy was conducted, and in 3 % of the studies it was not possible to ascertain who was in the therapy session (individual, couple, family). In fact, most did not state clearly who was in the therapy sessions, but usually the raters were able to derive this information from a description of the sort of therapy in the introduction to the study. However, this judgment reduces reliability, as was evident in the inter-rater agreements, which were slightly lower in these areas.

Discussion

This review was specifically designed to chart the full range of research methodologies currently being reported. Studies were classified into eight different methodologies but it is clear that those claimed by the authors to be RCTs dominate despite the ongoing debate in the field about the applicability of a medical model of research in psychotherapy and counseling (Duncan et al. 2009; Stratton and Harris 2012), Couple and Family Therapists are engaging with the dominant outcome research paradigm, and the field is generating RCTs in a broad range of clinical areas. Also, it was the RCTs that had the highest ratings of quality independently of the quality of randomization. Despite this higher rating, only 21 % of these RCTs in our review met criteria for a full description of sound randomization, and as these RCTs are generally on a limited scale, very few of them could be used in meta-analyses.

By the accepted standards of research publication, the articles reviewed present a strong case for the effectiveness

of CFT. As indicated in the introduction, both meta-analyses and less formal compilations of evidence have indicated effectiveness of CFT. Our data show that a substantial number of research studies applying different forms of therapy to patients with a wide range of conditions, researched using a great variety of methods including a rich array of outcome measures was published in sound journals and carried a consistent message of positive findings. We reported only the authors' own conclusions but each of these papers had been subjected to the standard scrutiny of refereed journals and the claims of the authors had been allowed to stand.

But there is a dilemma. Our exceptionally rigorous and detailed investigation of this corpus of research has shown substantial deficiencies in the reporting and a wide spread of quality in the methodologies. This does not cast doubt on the overall implications of this body of research but it does indicate that current standards of reporting mean that claims of effectiveness do not apply equally across the full range of clinical practice.

A broader set of issues arises from our finding that research is disproportionately available in areas of societal concern such as substance misuse, child/adolescent behavioural problems, and adult psychosis. This concentration leaves other areas of good and effective practice without research support and vulnerable to being denied to clients in systems of managed care or state provision. The long-term solution is for researcher interest and research funding to be directed to promising areas and forms of application of appropriate therapies. At present the restriction of recommended therapies to those supported by high quality randomized control trials can lead to grant bodies selectively funding forms of therapy that lend themselves to standardized application when applied to readily diagnosed 'mental illnesses'. Which can feed back to convince the bodies that manage therapy provision to conclude that only a narrow range of therapies have good enough evidence to be recommended.

A more immediate solution would be to change the criterion away from only sanctioning therapy for which there is convincing evidence of one particular kind, to only resisting therapies that research has shown not to be as effective as existing alternatives. That is, to accept the accumulating evidence that there are not major differences in effectiveness attributable to the mode of therapy when it has a sound rationale, rigorous training, and when the form of therapy makes a good fit with the style of the therapist: Hubble et al. (1999); Sexton (2007); Simon (2006); Sprenkle and Blow (2007). But this solution loops back to the problem of selective publication. It would rely on the published availability of research that indicated a lack of effectiveness of certain therapies. That is, researchers writing up, and journals publishing, negative results. Our analysis strongly suggests that this would require a substantial and well publicized change in journal policy worldwide.

Although the success of the field of CFT in generating enough high quality RCTs to be included in some international guidelines, the results of this survey point to a concern in relation to the comparison group that are used in the research. It is the policy of agencies such as the National Institute for Health and Clinical Excellence in the UK to make recommendations on the basis of difference from a comparison treatment regardless of the quality of that treatment. In a meta-analysis of evidence-based treatments for adult depression and anxiety, Wampold et al. (2011) reported that the treatments were more effective than treatment as usual (TAU) but only if the TAU was not another form of psychotherapy. Our finding that stronger claims were made for weaker comparison treatments is to be expected, but is relevant when treatments are judged by the size of difference from comparison regardless of what that comparison is. Better research (using a realistic comparison) is less likely to be able to claim effectiveness than research that compares with a poor alternative therapy or no therapy at all. Such a strategy is not a route to identifying the most effective therapies.

We noted various distortions that appear to have been introduced by the factors discussed above. Publishing strategies may lead authors to make the strongest claims they can for effectiveness of the therapy; journal preferences may operate synergistically with those of research funders to prioritize research in areas of societal concern; objective measures may be favored over alternatives that might more realistically reflect the objectives of CFT; the better established and therefore more acceptable measures are those that assess individuals rather than relationships. Such preferences are not arbitrary but perhaps could usefully be more explicitly recognized.

The limitations of this review arise from two sources the first of which is the papers selected for review. Our resources did not allow for translation of articles not in English and we cannot claim that our sampling is fully comprehensive. A different review might have taken a stricter criterion for 'CFT' but we were concerned to include all of the research that would be of value to CFT practitioners. Alternatively, some might also argue that all psychoeducation or all programs of parent training should be included. The second source of limitation arose from the limitations in what was reported. Authors have to decide where they will spend their 'allocation' of journal space and inevitably, some of the fine detail of the research may be omitted in favor of other aspects of the report. Very substantial efforts were expended on processing all of the content of the papers and where possible, following up

other reports of the research or studies by the same research group, to fill out the information. But there remained gaps, which are apparent in the reduced numbers, and the reduced reliability of coding, on some variables.

Conclusions

Within the limits of accepted standards of publication, the research supports the belief in the field that many different forms of CFT can be effectively applied in varied contexts for the benefit of people (though not evidencing the full range of diversity of people) struggling with a great variety of difficulties.

The published studies do not give confidence that ineffective therapy would be identified. Current concerns about differential effectiveness of therapists, or the possibility of ineffective approaches, are not addressed by the available research. There has to be a concern that negative results are excluded to the extent that repeated failures to demonstrate effectiveness of certain therapeutic approaches or practices could be present in research but will not be reported.

Journals have demonstrably not been constrained to the narrow requirements promulgated by those who want a research methodology that will provide a definitive answer to how to choose just one manualised therapy for each diagnostic category. But they have not either been able to require fully adequate levels of reporting.

This review has indicated a clear lack of CFT outcome research in some areas where it would be expected because of the nature of the presenting difficulties and the need for family members, and other systems to be involved in the individual's care. In particular there is very little research into CFT with older adults, with young people during the transition from adolescence to adulthood, and people of all ages with Learning Disabilities. Therapists are working in these areas in considerable numbers but without the benefit to their work that a sound research base could provide.

In 2002, Ogles, Lambert, and Fields, reviewing a range of psychological studies concluded that, "the simplicity and lack of precision in early studies has given way to such a great diversity and even chaos there is an obvious need for integration and organization"(p. 12). We hope that this review offers a fair representation of the diversity of research and practice in CFT and that it may contribute to the needed integration and organization.

Acknowledgments The authors wish to express their appreciation of the financial support and encouragement of the UK Association for Family Therapy (AFT) and the UK Council for Psychotherapy (UKCP), and to members of the research committees of both organisations who gave time and academic support to the survey.

Appendix

Strategy for Literature Searching

This section describes the search strategy adopted for searching the bibliographic databases. The search strategy was informed by the following guidelines: Centre for Reviews and Dissemination (2001). Undertaking Systematic Reviews of Research on Effectiveness. Centre for Reviews and Dissemination's Guidance for Those Carrying Out or Commissioning Reviews, CRD Report No. 4, 2nd edition. Centre for Reviews and Dissemination, York. Also, attendance of the Cochrane Systematic Review course held in Leeds in May 2006 by one of the researchers in the team (GP) helped to define the research question and develop the search strategy.

Initial "scoping" searches were carried out on the OVID databases, Medline and PsycInfo, using the following keywords: "family therapy", "couples therapy", and "marital therapy". The purpose of carrying out this exercise prior to developing the final search strategy was to compile a list of further keywords associated with CFT that may have been overlooked, and can thus be included in the final search strategy (on the complete reference of a particular article, Medline and PsycInfo provides other keywords and MeSH headings that the article has been indexed under). Doing this ensures a more accurate and comprehensive search strategy, that is less likely to overlook some potentially important references. This referencing had to be repeated for all the databases as they all operate different referencing systems.

In order to ensure that the final search strategy was as comprehensive and inclusive as possible, we employed a range of MeSH headings (category headings) and keywords (additional search terms to refine the search strategy, and ensure that no articles were overlooked), using Boolean Logic (AND/OR) to combine different components of the search strategy. Furthermore, it was known that English spellings in UK-published journals sometimes varied from the spellings in non-UK journals. To overcome this problem, wild cards (denoted by an asterisk, *) were used to capture spellings in both UK and non-UK journals, and truncation of keywords was used to capture text words with a common root. We also utilized adjacency operators in our search strategy.

The following bibliographic databases were searched for relevant papers: AMED; British Nursing Index and Archive; CINAHL; EMBASE; Global Health; MEDLINE; MEDLINE In-Process and Other Non-Indexed Citations; PsycINFO. Web of Science (2000 – 2009) and the Cochrane Library (2000–2009) were also searched.

A representative example of the search terms used is: 1—Family Therapy; 2—Relational Therapy; 3—Interpersonal

Therapy; 4—Biopsychosystem Therapy; 5—Family Therapy; 6—Marital Therapy; 7—Couples Therapy; 8—Systems Psychotherapy; 9—Narrative Therapy; 10—Family Psychotherapy; 11—Solution focused Therapy; 12—Family intervention; 13—Couples Psychotherapy; 14—Marriage Therapy; 15—(1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14); 16—limit 15 to (english language and year = "2000–2009").

The initial inclusion criteria were applied to the database search results; if it was not possible to judge from the title and abstract alone whether an article was relevant to our study, the full text was obtained to assess whether an article passed the initial inclusion criteria. The full text of all articles deemed relevant to our review judged by the initial inclusion criteria was obtained.

To increase the rigor of our search, we cross checked the existing search against the reference lists of relevant Cochrane and other reviews and selected included articles, to highlight any studies that we had not previously extracted. Further relevant studies were found from hand-searching relevant academic journals that had repeated citations. Primarily: The Journal of Family Therapy, the Journal of Marital and Family Therapy, the American Journal of Family Therapy, Family Process, and the Journal of Marriage and Family. Furthermore, using Web of Science, we performed Cited Ref searches, and searched for studies done by specific authors identified to be carrying out important research in this field, as well as searched existing studies in the Principal Investigator's personal reference collection. Finally, an interim version of the list of articles was sent to leading researchers in the field with a request to provide any missing references.

References

Carr, A. (2006). *Family therapy: Concepts, process and practice* (2nd ed.). Chichester: Wiley.

Carr, A. (2012). *Family therapy: Concepts, process and practice* (3rd ed.). Chichester: Wiley.

Carr, A. (2014a). The evidence-base for family therapy and systemic interventions for child-focused problems. *Journal of Family Therapy, 36*(2), 107–157.

Carr, A. (2014b). The evidence-base for couple therapy, family therapy and systemic interventions for adult-focused problems. *Journal of Family Therapy, 36*(2), 158–194.

Crane, D. R., & Payne, S. H. (2011). Individual and family therapy in managed care: Comparing the costs of treatment by the mental health professions. *Journal of Marital and Family Therapy, 37*(3), 273–289. doi:10.1111/j.1752-0606.2009.00170.x.

Crowther, M. A., & Cook, D. J. (2007). Trials and tribulations of systematic reviews and meta-analyses. *Hematology, 1*, 493–497.

Dallos, R., & Draper, R. (2010). *An introduction to family therapy*. New York: Open University Press.

Duncan, B. L., Miller, S. D., Wampold, B. E., & Hubble, M. A. (Eds.). (2009). *The heart and soul of change: What works in therapy* (2nd ed.). Washington DC: APA Press.

Froyd, J., Lambert, M., & Froyd, J. (1996). A review of practices of psychotherapy outcome measurement. *Journal of Mental Health, 5*, 11–15.

Hazelrigg, M. D., Cooper, H. M., & Borduin, C. M. (1987). Evaluating the effectiveness of family therapies: An integrative review and analysis. *Psychological Bulletin, 101*, 428–442.

Hubble, M., Duncan, B., & Miller, S. (1999). *The heart and soul of change: What works in therapy*. Washington DC: APA Books.

Lebow, J., & Gurman, A. (1995). Research assessing couple and family therapy. *Annual Review of Psychology, 46*, 27–57.

Nichols, M. P. (2012). *Family therapy: Concepts and methods* (10th ed.). Upper Saddle River: Pearson.

Norcross, J. C., & Beutler, L. E. (2011). Integrative psychotherapies. In R. J. Corsini & D. Wedding (Eds.), *Current psychotherapies* (9th ed.). Belmont, CA: Brooks/Cole Cengage.

Ogles, B. M., Lambert, M. J., & Fields, S. A. (2002). *Essentials of outcome assessment*. New York: Wiley.

Pote, H., Stratton, P., Cottrell, D., Shapiro, D., & Boston, P. (2003). Systemic family therapy can be manualised: Research process and findings. *Journal of Family Therapy, 25*, 236–262.

Retzlaff, R., von Sydow, K., Beher, S., Haun, M., & Schweitzer, J. (2013). The efficacy of systemic therapy for internalizing and other disorders of childhood and adolescence: A systematic review of 38 randomized trials. *Family Process, 52*, 619–652.

Rivett, M. (2008). Towards a metamorphosis: Current developments in the theory and practice of family therapy. *Child and Adolescent Mental Health, 13*, 102–106.

Sanderson, J., Kosutic, I., Garcia, M., Melendez, T., Donoghue, J., Perumbilly, S., et al. (2009). The measurement of outcome variables in Couple an Family Therapy research. *The American Journal of Family Therapy, 37*, 239–257.

Sexton, T. L. (2007). The therapist as a moderator and mediator in successful therapeutic change. *Journal of Family Therapy, 29*(2), 104–108.

Sexton, T. L., & Alexander, J. F. (2002). Family-Based Empirically Supported Interventions. *The Counseling Psychologist, 30*(2), 238–261.

Sexton, T. L., Datachi-Phillips, C., Evans, L. E., LaFollette, J., & Wright, L. (2013). The effectiveness of couple and family therapy interventions. In M. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behaviour change* (6th ed., pp. 587–640). New York: Wiley.

Shadish, W. R., & Baldwin, S. A. (2003). Meta-analysis of MFT interventions. *Journal of Marital and Family Therapy, 29*, 547–570.

Shadish, W. R., Montgomery, L., Wilson, P., Wilson, M., Bright, I., & Okwumabua, T. (1993). The effects of family and marital psychotherapies: A meta-analysis. *Journal of Marital and Family Therapy, 21*, 992–1002.

Simon, G. (2006). The heart of the matter: a proposal for placing the self of the therapist at the center of family therapy research and training. *Family Process, 45*(3), 331–344.

Sprenkle, D. (2012) Special issue of the Journal of Marital and Family Therapy on The effectiveness of Couple and Family Therapy, volume 38, no 1.

Sprenkle, D., & Blow, A. (2007). The role of the therapist as the bridge between common factors and therapeutic change: more complex than congruency with a worldview. *Journal of Family Therapy, 29*(2), 109–113.

Stratton, P (2011). *The evidence base of systemic family and couples therapy.* Association for Family Therapy, UK. www.aft.org.uk. Accessed on 2011/12/05.

Stratton, P., & Harris, T. (2012). Cultivating and customising the therapy relationship. *The Psychotherapist, 51*, 4–6.

Stratton, P., Reibstein, J., Lask, J., Singh, R., & Asen, E. (2011a). Competences and occupational standards for systemic family and couples therapy. *Journal of Family Therapy, 33*, 123–143.

Stratton, P., Silver, E., Nascimento, N., Powell, G., McDonnell, L., & Nowotny, E. (2011) *Review of family, couples and systemic therapy outcome research 2000–2009*. AFT, UK. http://www.aft.org.uk/SpringboardWebApp/userfiles/aft/file/Research/survey%20of%20outcome%20research.pdf. Accessed on 2014/08/08.

Vetere, A., & Dallos, R. (2003). *Working systemically with families*. London: Karnac.

von Sydow, K. (1999). Sexuality during pregnancy and after childbirth: A meta-content analysis of 59 studies. *Journal of Psychosomatic Research, 47*, 27–49.

von Sydow, K., Beher, S., Schweitzer, J., & Retzlaff, R. (2010). The efficacy of systemic therapy with adult patients: A meta-content analysis of 38 Randomized Controlled Trials. *Family Process, 49*, 457–485.

von Sydow, K., Retzlaff, R., Beher, S., Haun, M., & Schweitzer, J. (2013). The efficacy of systemic therapy for childhood and adolescent externalizing disorders: A systematic review of 47 RCTs. *Family Process, 52*, 576–618.

Wampold, B. E., Budge, S. L., Del Laskaa, K. M., Rea, A. C., Baardsetha, T. P., Flűckigera, C., et al. (2011). Evidence-based treatments for depression and anxiety versus treatment-as-usual: A meta-analysis of direct comparisons. *Clinical Psychology Review, 31*, 1304–1312.