Investigating couple relationships and change during couple and family therapy:

A specific focus on domestic violence and work functioning

Rune Zahl-Olsen

UiA University of Agder

Doctoral Dissertations at the University of Agder 266

Rune Zahl-Olsen

Investigating couple relationships and change during couple and family therapy:

A specific focus on domestic violence and work functioning

Dissertation for the degree of philosophiae doctor (ph.d.)

University of Agder
Faculty of Health and Sport Sciences

Sørlandet Hospital
Clinic for Mental Health
Department of Child and Adolescent Mental Health

2020

Doctoral dissertation at University of Agder 266

ISSN: 1504-9272
ISBN: 978-82-7117-967-0

Print: 07 Media, Kristiansand

Summary

The aims of this thesis were to investigate couple relationships and change in those relationships during couple and family therapy, with a specific focus on domestic violence and work functioning.

The first subsidiary aim was to assess, theoretically and empirically, what level of simplification should be applied to the comprehensive measure used in this study. The second subsidiary aim was to examine the physical couple and family violence among clients receiving couple and family therapy. The final subsidiary aim was to analyze the clients' change in work functioning from pre- to post-treatment.

The clients who participated in this study (N=841) can be considered as representative of clients in couple and family therapy in Norway when considering income, the number of children, educational level, prior experience with therapy, and levels of distress on individual and relational measures.

Domestic violence is commonly mentioned in the news because the consequences sometimes are fatal. However, research on domestic violence on clinical samples is sparse, and it has often failed to differentiate between emotional and physical violence. Furthermore, violence between others in the family besides the couple has, to the best of my knowledge, not been included. Outcome studies of couple and family therapy most often focus on relational functioning or individual distress. However, work functioning has, to the best of my knowledge, not been used as an outcome measure.

Paper 1 investigated the level of simplification that should be applied to the comprehensive measure used in this study, the Systemic Therapy Inventory of Change (STIC). Since the use of standardized measures implies simplification of what can be captured of human experiences, the challenge of such use is to ensure that the information collected represents the human experience as close as possible to reality. However, the measure should be feasible when considering the time and energy spent in answering and examining the results. Therapists and researchers need simplified information to get an overview of the complex lives of the clients, but if the overview is too simplistic, it might mislead more than inform. The question is, therefore, how much simplification can be conducted before the measures no longer inform us as therapists and researchers in a

credible way about the clients' levels of distress and the process of change in therapy? Paper 1 investigated what level of simplification could be conducted on one of the scales of the comprehensive self-assessment questionnaire, STIC. We found that the scale could not be simplified beyond the sub-scales model and still be clinically and scientifically useful to understand the clients and the possible change in their lives. The conclusion was that even though it is possible to simplify conceptually different items into one total scale, this approach should not be taken when evaluating problems or changes in therapy. The reason is that the simplification could increase the risk of misunderstanding clients and their therapeutic experience. In this thesis, I discuss the result from the analysis concerning measuring changes, in general, and specifically in couple and family therapy when using comprehensive measures.

Paper 2 examined physical couple and family violence among clients seeking therapy with a specific focus on identifiers and predictors of the violence exerted. In total, 25 % of our sample, derived from a clinical setting, responded that they had experienced physical family violence, on a range from "sometimes" to "all of the time." Family violence was detected in 50 % of the cases with violent couples, compared to 18 % in the cases with non-violent couples. A model for predicting physical violence was tested. The model explained 53 % of the physical violence, and the strongest predictor was found to be the expectation level towards the partner, specifically regarding household chores. This fact means that the more the clients experienced their partners expecting too much of them, the more physical violence they reported. The more they experienced being filled with anger towards their partner, the more physical violence they reported happening. Self-control was found to be a negative predictor of physical violence, meaning that more self-control was associated with less physical violence. Contrary to our expectations, the more sexually satisfied they reported being in the relationship, the more physical violence they experienced.

Paper 3 analyzed the change in work functioning from pre- to post-treatment, concerning distress on three dimensions. In this study, the clients reported their individual level of distress as well as levels of distress on two relational dimensions. At the individual level, we measured the level of depression with the Becks Depression Index (BDI). The distress in their couple relationship was measured with the Revised Dyadic Adjustment Scale (RDAS), and family functioning was measured with the Family Assessment Device (FAD). The group

of clients improved from start to end of treatment on work functioning as well as on the individual-, couple- and family measures. The level of work functioning was significantly better predicted when the relational measures were included, compared to when only the individual measure was used. None of the levels of these measures (i.e., individual, couple, or family measures) at pre-treatment could predict the work functioning at post-treatment. However, the change in these measures from pre- to post-treatment could predict 54 % of the work functioning at post-treatment when controlling for the level of work functioning at pre-treatment.

List of papers

1. Zahl-Olsen, R., Gausel, N., Håland, Å.T., Tilden, T. (In review). Monitoring therapeutic change through diversity or simplicity? A conservative, critical Confirmatory Factor Analysis test of a Routine Outcome Monitoring system.

2. Zahl-Olsen, R., Gausel, N., Zahl-Olsen A., Bjerregaard Bertelsen T., Håland Å.T., Tilden T. (2019). Physical couple and family violence among clients seeking therapy: identifiers and predictors. *Frontiers in Psychology, section Psychology for Clinical Settings.*

3. Zahl-Olsen, R., Håland, Å.T., Gausel, N., Wampold, B., Tilden, T. (2019). Change in work functioning from pre- to post-treatment in feedback-informed couple and family therapy in Norway. *Journal of Family Therapy*.

Acknowledgments

It is hard to define an actual starting point of this thesis, but I recall when PhD Åshild Tellefsen Håland, one of the supervisors, came to my office just after I had finished a therapy session, asking if I wanted to participate in a research project. I was skeptical about the quantitative approach and the online feedback system that we were about to try out in clinical practice. Initially, I did not think it could gain any useful information for the therapy my colleagues and I performed. However, Åshild convinced me that it would be interesting, even fun, to be a part of the research project. As time passed by, I found the data we collected to be interesting and even helpful in therapy. At the same time, I slowly understood more and more of the research aspects of the project and started to ask questions about the data. Together with some people from the local team consisting of therapists and researchers, I twice visited our collaborators in the US and participated in discussions about the study design and how to understand the preliminary results. I wrote several drafts of research projects before I finally submitted and got a grant from Sørlandet Kompetansefond for the present project.

I wish to express my most sincere gratitude to my main supervisor, Terje Tilden. He was the initiator of the STIC project in Norway, is still leading the project, and has tirelessly answered emails, read drafts and made comments, discussed on Skype and telephone, and invited me to the research department at Modum Bad several times to keep the momentum in this study. His patience, thorough comments on drafts, and supportive approach have been invaluable.

I also want to express my gratitude to my co-supervisors, Åshild Tellefsen Håland and Nicolay Gausel. Tellefsen Håland has, for the last two years, been the leader of the research department at the department for Child and Adolescent Mental Health (ABUP), Sørlandet Hospital, where I am working. She has been supporting this project in several ways, as part of the steering group of STIC in Norway, local administrator for the first years of the project, and one of the co-supervisors for this project. She has provided, as my leader, the needed support but also been the one following the project on a day to day basis, always available for questions and suggestions on how to move on.

Nicolay Gausel has brought statistical knowledge and feedback from an adjacent field that has lifted the discussions and improved the project and the papers in a

way not possible without his participation. At critical time points, he invited me to his office to do analyses under his supervision.

Please accept my deepest gratitude.

I also want to credit all the coauthors for their important contributions to the papers.

Thanks to Sørlandets Kompetansefond, Sparebanken Sør, and department of child and adolescent mental health (ABUP) that has financially supported this project.

Thanks to my family, Agnes, Marius, Jonas and Julianne, for your support and understanding during these years.

Finally, I am deeply grateful to God who gave me life, joy, energy and perseverance to fulfill this project.

Table of contents

List of abbreviations in this thesis

APA	American Psychological Association
BDI	Beck Depression Inventory
CFT	Couple and Family Therapy
DV	Domestic Violence
EBP	Evidence-Based Practice
EBT	Evidence-Based Treatment
FAD	Family Assessment Device
IPS	Individual Problems and Strength scale
IPV	Intimate Partner Violence
RCT	Randomized clinical trial
RDAS	Revised Dyadic Adjustment Scale
ROM	Routine Outcome Monitoring systems
STIC	Systemic Therapy Inventory of Change

List of tables and figures in this thesis

1 Introduction

1.1 Aims of this thesis

The aims of this thesis were to investigate couple relationships and change in those relationships during couple and family therapy (CFT), with a specific focus on domestic violence and work functioning.

The first subsidiary aim was to assess, theoretically and empirically, what level of simplification should be applied to the comprehensive measure used in this study. The second subsidiary aim was to examine physical couple and family violence among clients receiving CFT. The final subsidiary aim was to analyze the change in the clients' work functioning from pre- to post-treatment.

In the following introduction, I will start by placing this thesis within its context. I start by defining CFT and describe its historical background and philosophical basis. I will then present some main developments and findings within CFT research. Further, I will present a more specific background for each of the three papers. I start by presenting some of the complexity of measuring change in psychotherapy and, thereby, laying out the rationale for Paper 1. I will then present the different focuses that have been applied within psychotherapy research and present arguments for why I have chosen to focus on domestic violence (Paper 2) and work functioning (Paper 3) in this thesis.

1.2 Definition and history of Couple and Family Therapy

1.2.1 Definition of CFT

The term CFT captures a vast area of different perspectives and treatment programs (Lorås & Ness, 2019; Ness, 2017; Sexton, Datchi, Evans, LaFollette, & Wright, 2013), even if all are related to general systems theory (Von Bertalanffy, 1972) and communications analysis (Bateson, Jackson, Haley, & Weakland, 1956). CFT is the branch of psychotherapy with a focus on working with families and couples in intimate relationships where the aim is to contribute to change and development (Johnsen & Torsteinsson, 2012). The preferred focus of change is on the relational level rather than on the individual, i.e., the couple or the family. Change is, from this perspective, most often seen in patterns of cooperation or how people relate to each other (Johnsen & Torsteinsson, 2012). CFT is

distinguished from individual treatment when it comes to the focus of the treatment sessions, and the number of clients showing up to therapy sessions. In CFT, most often, a couple or a family meet jointly for therapy. CFT is not individual therapy with listeners, but a therapy where all are participants and all are viewed as being part of the problem for which they are seeking help (de Flon, 2019). The therapeutic focus is, more often than not, interpersonal, such as on the relationships in-between the clients, rather than on intra-psychic topics within each individual client (Johnsen & Torsteinsson, 2012; Ness, 2017). Despite this main relational approach to CFT treatments, e.g., by focusing on couple satisfaction, psychosexual aspects, gridlocked communication, and family function (Sexton et al., 2013), there are CFT treatments targeting individual aspects such as anxiety, depression, enuresis, anorexia (Carr, 2014a, 2014b) where individual approaches are seen as beneficial within a relational treatment context. A CFT therapist who works from a systemic perspective does not neglect the importance of individual psychological processes but prefers to understand individual behavior in the context of the client's primary intimate relations (Hårtveit & Jensen, 2004). This thesis is investigating clients participating in CFT treatment that has this relational focus.

1.2.2 Historical background for CFT

There has been a change in perspectives from the early years of CFT until now. The field, especially here in Norway, has been dominated by a social-constructionist perspective emphasizing subjective experiences and is skeptical about the usefulness of knowledge gained from research on groups of clients (Hårtveit & Jensen, 2004) as this thesis is. However, that skepticism was not the case in the early years of CFT, and it does not dominate the CFT field so much anymore (Johnsen & Torsteinsson, 2012; Lorås & Ness, 2019; Ness, 2017).

The development of CFT as a therapeutic intervention was inspired by two major influences. The first influence came from systems theory, cybernetics, and biology and the other from psychoanalytically oriented therapists who were dissatisfied with the progress in treating individual clients and, therefore, started to include family members in therapy (Johnsen & Torsteinsson, 2012). As described by Johnsen and Torsteinsson (2012), the early developers of CFT treatments were all professionals with strong connections to academic institutions where quantitative research methods were a natural part of thinking. Hence,

Johnsen and Torsteinsson (2012) conclude that research has been a natural part of CFT since the beginning. CFT developed with a patchwork of influences from psychotherapy but also other fields, such as systems theory, cybernetics, mathematics, and biology (Tilden, 2010). In the early stages, the field of CFT was strongly influenced by modernism, and at that stage of CFT, the therapist was viewed as an expert that could assess problems within the system and intervene to change those (de Flon, 2019). Later, when the therapist was viewed as a part of the system, contrary to being just an outside observer, the field of CFT abandoned the view of the therapist as an expert (de Flon, 2019). CFT started with a modernistic perspective, moved through the constructivist perspective to the social-constructionist perspective that is dominating today (de Flon, 2019), a perspective based on post-modernism and post-structuralism. From a modernistic perspective, a therapist would try to understand the patterns of interactions within a family by investigating parts of this picture by splitting it up and scrutinize those parts (e.g., communication patterns) (Hårtveit & Jensen, 2004). Within the constructivist perspective, the main idea is that there is no objective truth, only subjective experiences (Von Glasersfeld, 1988). The social-constructionist perspective is a reaction to the constructivist perspective and highlights the importance of interaction between people when meaning is created (Gergen, 2015).

In contrast to this perspective, Pinsof (1992), a CFT therapist and researcher, introduced *interactive constructivism* that views knowledge as progressive. The author describes that science cannot confirm reality or truth, but that it is best at telling us that something is not true - that our construction or hypothesis is wrong. Other theories have also been used among CFT researchers; one example is critical realism (Collier, 1994) used by Solem, Thuen, and Tilden (2008). This theory makes a clear distinction between what is real and our assumptions about reality (Solem et al., 2008). It is in an environment of these influences that this thesis is written. To understand how these developments in the field impact this thesis, I will, in the next chapter, describe the development of the philosophy of science that followed the development of CFT.

1.2.3 The philosophy of science within CFT

As described, there has been a development of the theories influencing CFT from the early start in the sixties until now. These changes have severely influenced

ontological and epistemological aspects of CFT theory and practice. *Ontology* is about what we can obtain knowledge about, and when it comes to therapy and research on therapy, specifically the question: What knowledge is possible to obtain access to that is useful for understanding the clients and the process in therapy (Willig, 2019)?

Since several of the CFT approaches are inspired by post-modern philosophy (Johnsen & Torsteinsson, 2012), there is an ontological debate on whether it is possible to get access to knowledge useful for a specific individual, couple or family meeting a therapist by the use of information gained from other clients, especially groups of clients who met other therapists (i.e., nomothetic knowledge). The standardization of treatments that forms the basis for much of the quantitative studies is, for many therapists, seen as ignoring the complexity of practice (Johnsen & Torsteinsson, 2012). Further, there has been skepticism within the systemic CFT field because empirical knowledge is associated with experimental quantitative trials in laboratory-like settings that CFT therapists do not find relevant for their clinical context (Pinsof & Wynne, 2000). Tilden (2010) argues that this, in particular, implies knowledge gained through RCT's that use standardized treatment manuals and quantitative questionnaires. Indeed, this is in strong contrast to a CFT practice full of unique processes, that some experts suggest are better captured by literature and poetry than by a reductionistic approach (Johnsen & Torsteinsson, 2012). Angier (2007) describes this reductionism used in science as "some picking apart and focusing on one or two variables at a time" (p.192). I will, in the following, use the terms reductionism and simplification interchangeably.

The closely related *epistemological* debate that follows mentions the ontological debate about if it is possible to obtain clinically useful knowledge. How should this knowledge be attained? Kazdin (2003) writes: "The ways in which psychotherapy is studied depart considerably from how treatment is implemented in clinical practice. Consequently, the extent to which findings can be applied to work in clinical settings can be challenged" (p. 262). Kazdin here presents a dilemma on how to conduct research in a way that will bring clinically useful information. Quantitative research is often seeking to simplify and control for factors that might contaminate the results, while the clinical practice is full of complexity and comorbidities (Johnsen & Torsteinsson, 2012).

Within epistemology, there is an important differentiation between objectivity and subjectivity (Johnsen & Torsteinsson, 2012). Within objectivity, one has the assumption that what is observed by following certain methods is objectively true, i.e., true to everyone. Since CFT theory is strongly influenced by social-constructionist ideas (de Flon, 2019), objective knowledge is questioned. Subjectivism means that truth should start with the subject, not the object. The person's experiences, feelings, attitudes, affiliations, and the context of the person is the basis of knowledge, not any objective, measurable aspect. This approach implies that what is true to you does not have to be true to me (Gergen, 2015). In other words, if we on an ontological level conclude that it is possible to obtain useful knowledge applicable for CFT settings, from an epistemological level, it is hard, if not impossible, to attain that knowledge.

All these ontological and epistemological questions are important to raise as this thesis, based on quantitative data, is conducted within CFT that often advocate qualitative designs and even more so as it includes data from a randomized clinical trial (RCT) study. CFT therapists, within the language-systemic tradition, argue that the standardization of treatments that are required for RCT's, cannot be the most important part of preferred practice (Johnsen & Torsteinsson, 2012). Some even argue that family therapy, due to its theory drivenness, is not particularly interested in what the research says about effective treatments (Johnsen & Torsteinsson, 2012). Campbell (2003) argues that it is particularly true for CFT therapists because they find little value in information generated by RCTs. However, as Larner (2004) describes, "there is an ongoing controversy about the political economy of evidence, how it is defined and who defines it. This question is not about evidence *or* no evidence but who controls the definition of evidence and which kind is acceptable to whom" (p. 20). Concerning this controversy, I find it important to mention the misconception that nomothetic knowledge, i.e., quantitative research, tries to apply to all that is only found to be true to the mean of the population studied (further to be discussed in chapter 7.3).

As described, there are several different perspectives within CFT. The perspective used in this study is within the integrative perspective (Pinsof, Breunlin, Chambers, Solomon, & Russell, 2015a) that derives from the before

mentioned theory of integrative constructivism (Pinsof, 1992). Applying the integrative perspective means that I see the individual client as the expert of his/her life (idiographic knowledge), but that information from other clients, in other settings (i.e., nomothetic knowledge), might be useful. Information originating from other settings is viewed as something that can help to inform, to understand, or contrast, but also as a tool for rich and meaningful conversations with the clients (Olkowska, Sundet, & Karlsson, 2018).

1.3 Developments and current state of CFT research

This thesis is answering to a specific field, CFT research. Research within this field has, from the start, been conducted from a variety of theoretical perspectives and has thus been a part of developing CFT treatments (Johnsen & Torsteinsson, 2012). One early example is from Maudsley (Eisler et al., 2000; Morgan & Russell, 1975; Russell, Szmukler, Dare, & Eisler, 1987), where they assessed and developed their CFT treatment for anorexia nervosa and bulimia nervosa through the use of RCTs. They found the results useful for their clinical practice and as new questions arose, they applied research to investigate and to gain understanding. Sometimes the results challenged the therapists to be more flexible and accept interventions that they, in the beginning, were not comfortable with, however, the adjustments were found to be effective (Johnsen & Torsteinsson, 2012).

The part of the research field that adheres to the quantitative collection of data has now strong evidence that CFT treatment is effective (Sexton et al., 2013). Carr (2014b) concludes in his review that 71 % of the families receiving family therapy are fared better after therapy and follow up than families in control groups. Several meta-analyses have been conducted (see, e.g., Powers, Vedel, & Emmelkamp, 2008; Shadish & Baldwin, 2002) and when Crane and Christenson (2012) summarized the cost-effectiveness studies within CFT, they found that family therapy is associated with greater benefits than individual therapy. However, many of the different CFT practices have not been investigated to assess efficacy or effectiveness (Gurman, Kniskern, & Pinsof, 1986). Indeed Sexton, Alexander, and Mease (2004) conclude in their review of CFT that there is a lack of research concerning the majority of the contemporary CFT

approaches. Sexton and colleagues (2013) describe development in CFT research "from reductionistic approaches that simplify complex clinical phenomena and moved toward research methods capable of capturing the unique interplay between client-presenting problems, therapeutic factors, demographic diversity, and model-specific process to outcome variation" (p. 587). Even so, Sexton and colleagues (2013) found that the vast majority (85%) of family therapy studies were outcome studies, and few (15%) were investigating the process. They also conclude that the manual based CFT (e.g., Multisystemic therapy and Cognitive-behavioral family therapy) treatments are dominating the CFT research field.

However, as Sexton and colleagues (2013) point out, more research is needed. Notably, they mention broadening the range of clinical problems studied, expanding the range of treatment programs, and enlarging CFT research beyond the study of outcomes. They also suggest more comparative effectiveness trials and the use of diverse research methods. This study mainly answers to the first of those suggestions as it in Paper 2, investigates physical violence and in Paper 3 investigates work functioning.

At the current stage, we can conclude that CFT works, but we still know relatively little about *why* the interventions work (Sexton et al., 2013). Sexton and colleagues (2013) also state that most of the research focuses on "specific and comprehensive programs rather than the general approaches and individual techniques that are often the intervention of choice in clinical practice" (p. 628). Furthermore, most studies are conducted within university settings with young students as participants (Sexton et al., 2013) and, therefore, not representative of the population of clients seeking help at CFT clinics. As a response to these shortcomings, this thesis is [a)] not focusing on any specific manualized treatment intervention, but on general practice in a [b)] public setting and [c)] not with particularly young clients.

Since I am a Norwegian researcher researching a sample of clients from our country, I find it important to place this thesis also in the context of Norwegian CFT research. Research within this field has been sparse in Norway, mainly applying qualitative designs (see, e.g., Helgeland, 2014; Jensen, 2006; Loras, 2016; Ness, 2011; Oanes, 2016; Sundet, 2009), however with a recent increase with both qualitative and quantitative approaches, e.g., Anker, Duncan, Sparks,

and La Greca (2009); Anker, Owen, Duncan, and Sparks (2010); Sparks (2015); Tilden (2010); Tilden et al. (2019b). The doctoral theses of Sundet (2009) and Oanes (2016) investigate the use of Routine Outcome Monitoring systems (ROM) within CFT in Norway from a qualitative perspective and are thus related to some of what has been investigated in this study. Indeed, I could have implemented qualitative approaches in this study. Concerning Paper 1, I could have interviewed therapists and clients investigating their perspective on the measure for a more in-depth knowledge than the statistical approach I used. However, Zahl-Olsen and Oanes (2017) already investigated both the therapists' and clients' experiences of the use of that measure. More important, the decisions on what methods to use in this study was defined at the time I came into the project (This point is further discussed in 7.2.1).

1.4 Dilemmas when using questionnaires to measure change

Applying standardized, quantitative measures to tap all areas of human experience relevant for psychotherapy is a hard task (Thagaard, 2003). Hence, in its nature, the choice of applying quantitative approaches as aids in clinical practice has an embedded limitation, because every single measurement focuses on a limited part of human experience. As such, this approach is reductionistic as it is not able to capture the bigger picture of, e.g., interactions and circularity (Punch, 2014; Teddlie & Tashakkori, 2009). For instance, some of these measures were designed to capture only specific areas of people's lives, e.g., communication skills (Takahashi, Tanaka, & Miyaoka, 2006). Other measures were designed with a more general approach, e.g., the experience of life in general (Loge & Kaasa, 1998). As a consequence of the democratization of psychotherapy that was part of the post-modern influence, user involvement and empowerment have been regarded as central values impacting our field (Tilden, 2017).

One means to realize the intent of democratization was to make sure that the voice of the client was realized as part of clinical practice (Tilden, 2017). For instance, performed by collaboration between client and therapist on defining aims for the outcome, by assessing the course of treatment, and by choosing the therapeutic approaches as preferred by the client (Tilden, 2017). This approach implies that power in therapy has to a large extent been moved from the therapist to the client, representing a power shift that was addressed by the American

Psychology Association (APA) as they upgraded the client's knowledge regarding the knowledge of the experts, i.e., therapists and researchers, in what APA labeled Evidence-Based Practice (EBP) (APA-Task-Force, 2006). EBP will be discussed in chapter 2.4.4, but I will here point out that it lays a theoretical foundation for the importance of the use of the client's voice in therapy.

During the last twenty years, systematic and frequent use of client responses on standardized measures, ROM; (Lambert, 2010; Ogles, 2013; Tilden & Wampold, 2017) has evolved to assess better the change that happens during treatment. The aim of the use of ROM is to capture real-time information from the clients to inform the therapist reliably about the process and progress in therapy. In particular, ROM's ability to yield the therapist needed warning signals if the treatment is not heading in a desirable direction has been emphasized (Lambert & Shimokawa, 2011). ROM will be discussed in more detail in chapter 2.3.1, but I mention it here because a comprehensive ROM system, the Systemic Therapy Inventory of Change (STIC) (described in chapter 2.3.2), was used in this study.

However, quantitative measures, that forms the basis of any ROM, is reductionistic (Angier, 2007). The developers of these systems have chosen areas in the clients' lives they believe are of importance to assess. The clients are asked a set of questions within each of these areas, and before their answers are presented to the therapist, they are combined, i.e., simplified. The question we asked in Paper 1 was how much simplification could be conducted before the ROM systems contribute to confusing therapists and have those to misunderstand their clients rather than enlightening them as intended? The results and the discussions in Paper 1 are contributions to the debate of which measures to use, how much simplification can be applied, and how to develop new and better ROM measures. The results from Paper 1 also guided me in what measures to include or exclude in our search for answers in papers 2 and 3.

1.5 Should the focus be individual, relational, or societal?

Most research within the field of psychotherapy focuses on individual aspects, e.g., symptoms of depression and anxiety (Lambert, 2013a). Within the narrower field of CFT research, relational aspects have also been in focus, e.g., couple distress and family functioning (Sexton et al., 2013). However, societal aspects (e.g., work functioning and sick leave) have been outside the main focus in CFT research from the early start, until now (Gurman, Kniskern, & Pinsof, 1986;

Sexton et al., 2013). Since one of the basic theories of CFT is that any action or individual has to be understood not alone, but within its context and relations (Bateson, 1972) this thesis focuses on investigating couple relationships within the context of their family and society, still keeping in mind the individual who is a part of those relationships. Paper 2 focuses on one aspect of the individual client in the context of his or her couple and family relationships, i.e., domestic violence[1]. Paper 3 focuses on one aspect of the individual client's life in the context of society, namely work functioning, and how this is related to their individual level of depression, couple relationship and family functioning. Hence, individual, relational, and societal aspects are incorporated in this study.

1.5.1 Relational focus – Domestic violence

One relational aspect is domestic violence, and among the conducted research on this topic, the samples have mainly been drawn from a general population, not from a clinical population within the context of CFT (Jose & O'Leary, 2009). Out of the 262 studies in the latest review of Sexton and colleagues (2013), they only identified four studies where abuse was assessed. Since CFT treatment often deals with domestic violence, I find it surprising that the CFT research field has not investigated this in more detail. Domestic violence has got much attention in national media in Norway lately, as it often is related to fatal consequences for adults as well as for children. Indeed, on the day I am writing this in July 2019, one of the news headlines is about a mother who was sentenced to 21 years of detention for killing both her father and a former cohabitant.

Fortunately, outcomes from domestic violence are most often less severe. To the best of my knowledge, the few quantitative studies that exist within CFT on domestic violence are all from the US, and they do not differentiate between psychological, physical, and sexual violence (see Jose and O'Leary, 2009 for a review). Those studies also fail to include violence among others than the adult couple, even if it is known that when violence is occurring between the parents, the children are often involved (Appel & Holden, 1998; Fusco & Fantuzzo, 2009). Consequently, in Paper 2, the focus was to investigate physical violence

[1] The term domestic violence includes physical, emotional and sexual violence exerted between the couple and within the family.

and differentiate between violence that occurs between the adult couple and violence that includes others in their family.

The predictive model presented in Paper 2 (see appendix, Paper 2 page 19) is aimed to assist therapists in detecting physical violence between the couple and/or others in the family, even when it is not mentioned explicitly by the clients. How can a predictive model from quantitative research work as a clinical aid for the therapist? Predictors of violence identified by research can aid as signposts for therapists that violence might occur. For example, as identified in Paper 2, a high level of expectation towards each other when it comes to household chores was a predictor of physical violence. Hence, if high levels of expectation become an issue in therapy, the therapist could use this as a signpost that physical violence might be an issue. However, some therapists are skeptical about the use of knowledge gained through generalizations and categorizing clients from quantitative research creating group-based (nomothetic) knowledge (Lambert, 2013b). This situation applies especially to therapists practicing within the paradigm of social constructivism (Gurman et al., 1986) such as collaborative-dialogical practice (Anderson, 2019). These therapists claim that knowledge about the client always is local, individual, and context-dependent; hence, they rather embrace qualitative research designs that yield idiographic knowledge. Even if those therapists might have resistance in making use of nomothetic knowledge, it is suggested they can apply it to inform their questions in their collaboration with the clients to build understanding (Olkowska et al., 2018). Paper 2 is, therefore, a contribution on the road to document the prevalence of domestic violence among CFT samples and aid to clinically discovering and stopping damaging violent acts among clients participating in CFT treatment.

1.5.2 Societal focus – Work functioning

In Paper 3, this thesis incorporates the societal aspect as it investigates the change in work functioning from pre to post-treatment. This outcome is not identified in any of the 262 CFT studies investigated by Sexton and colleagues (2013) and to the best of my knowledge not used by any other study in CFT, even if work functioning is of great importance for the individual, their family, and society. Whisman and Uebelacker (2006) found that individual and relational distress is related to how well a person functions in a work setting. Indeed, more

than 50 % of sick leave is due to mental issues rather than physical medical reasons (Aronsson, Johansen, Marklund, Rønning, & Solheim, 2015). Moreover, the portion of those who go on sick leave due to mental health issues has been in a steady increase in recent years (Hensing, Andersson, & Brage, 2006). Since CFT treatment is found to be an effective treatment for both individual and relational distress (Carr, 2014a; Sexton et al., 2013) it is likely impacting the clients' ability to function at work. However, as mentioned, few studies within CFT have a societal focus (Sexton et al., 2013), and to the best of my knowledge, no study within CFT has investigated work functioning. Hence, in Paper 3, I investigate change on work functioning from pre- to post-treatment concerning individual and relational distress.

2 How to measure change in CFT?

As mentioned, it is not the case that all CFT therapists and researchers are totally emerged in the social-constructionist perspective and thereby negative to the usefulness of nomothetic knowledge (Johnsen & Torsteinsson, 2012). The field consists of a spectrum from those who argue for the ultimate importance of the use of standardized measurements, before you even meet the clients (Gottman & Gottman, 2015), to the ones who argue that the therapist should be as *not knowing* as possible and that predefined questions, therefore, are harmful (Anderson, 2019). If we are to define the right side of the spectrum to be the positive to standardized measures and the left to be negative, I would place those trained in Gottman therapy (Gottman & Gottman, 2015), Cognitive therapy (Tilden, 2019) and Integrative therapy (Gurman, Lebow, & Snyder, 2015) at the right-hand side. Therapists inspired by these approaches are used to standardized questionnaires and find them helpful in their CFT therapy. On the left-hand side, I would place practices like cooperative – dialogical therapy (Anderson, 2019) and narrative therapy (Maddigan & Nylund, 2019). For those on the left-hand-side, Paper 1 can meet some of the critiques of the usefulness of standardized measures. For those on the right-hand-side, Paper 1 contributes to the discussion of what measures to use. It investigated what level of simplification can be applied to the already reductionistic questions asked the clients for it to be useful and helpful for the therapist and researcher.

As previously mentioned, CFT is a diverse set of treatments focusing on no single agreed-upon approach nor outcome (Lorås & Ness, 2019; Ness, 2017; Sexton et al., 2013). For this reason, the CFT clinicians and researchers have not yet concluded on which outcomes are appropriate to apply in clinical practice and research (for a review, see Sexton et al., 2013). Hence, when measuring the outcome, there is, therefore, no single outcome to assess. It partly depends on the treatment given and the goal of the clients. I will, in the following chapter, present several aspects that need to be assessed before choosing outcome measures.

2.1 What to measure?

All since 1971, there has in psychotherapy research been a call for a core battery of quantitative assessment procedures and measures (Ogles, 2013), but it seems

as if researchers and developers of treatments continue to prefer using their own set of measures (Ogles, 2013). One plausible reason is that it is not possible for one measure to capture all aspects researchers would like to investigate. Indeed, Ogles (2013) found that most measures were only used once. Another reason for developing new measures might be fame and money. If a researcher develops a good measure that will be used by many, his name will be referred to in all papers using it, and if he additionally connects royalty to the use of the measure, it can be good business.

Maybe the most important argument is that those developing new treatment interventions often argue that they need an instrument that emphasize the central elements in their intervention (See, e.g., Comer and Kendall, 2013 and Wampold and Imel, 2015). One reason for this is to measure competency and adherence to the treatment model, which is inevitable to assure that the method really has been used.

Further, it is found that the more specific the measure is connected to the method, the more chance that the research will find that the method is effective (Wampold & Imel, 2015). However, there is a risk for allegiance where the researchers are also developers of the methods and also of the measures (Comer & Kendall, 2013). Researcher allegiance has been found to increase the effects of the treatment interventions, and results from meta-analyses investigating allegiance have found allegiance effects ranged up to 0.65 – a large effect (Wampold & Imel, 2015). Indeed, Wampold and Imel (2015) identified some studies where the researcher allegiance doubled the effect of the treatment compared to where the developer was not a part of the research team.

Of obvious reasons, no single measure can capture all human experiences, beliefs, emotions, thoughts, etc.; hence, a *multi-measurement strategy* is one possible solution. That means using more than one measurement to tap different aspects of human experience. For instance, measuring depression on an individual level and couple satisfaction on a relational level as we did in Paper 3. We used several standardized measures found to be valid and reliable. We used one measure for depression, one for couple relationship, one for family functioning, and one for work functioning (those will be described in detail in the method section). However, the measure for work functioning that we used in Paper 3 had never been used as an outcome measure within CFT before, even though the measure was thoroughly validated through other studies. When using

a multi-measurement strategy, there is a cost, most often paid by the clients answering more questions.

2.1.1 Selecting focus: Individual, relational, or societal?

There are different foci measures can have. The most common focus – because individual therapy is the majority of treatment modalities - is on the individual (Ogles, 2013). One example is the individual level of depression, as we used in Paper 3. Further, a measure may also target relational issues. This situation is often the case when measuring the level of distress and change in CFT (Sexton et al., 2013). The instruments applied in CFT mainly relate to these fields' theoretical and clinical objectives targeting relational issues, e.g., family functioning, dyadic satisfaction, problem-solving skills, and sexual satisfaction (Sexton et al., 2013). From a societal point of view, also measures of absenteeism from work, use of health services, and work performance are of high interest. In that respect, studies of cost-effectiveness (Crane, 2011; Crane & Christenson, 2012; Crane & Payne, 2011) are also highly interesting.

As mentioned in the introduction, this study incorporates individual, relational, and societal measures. However, an essential question to address is whether a measure measures what it is supposed to measure. Usually, that question is handled by advanced methods of analysis, such as confirmatory factor analysis (CFA). In a CFA, the factor loadings are observed, and items that load strongly enough are included in the scales, and weaker items are omitted (Kline, 2016). However, the constructs we measure in psychotherapy are, for the most part, theoretical constructs that cannot directly be measured (Field, 2018). Thus, a clear definition of what a good couple relationship is that all accepting might be hard to find.

Nevertheless, when using measures, these should be tested to see if they tap what they claim to measure on the sample of clients we are investigating. It should also be assessed if the measures make sense to clinicians and clients. Paper 1 considers all these aspects as it critically examines one of the scales in the STIC system from both a theoretical and empirical point of view.

2.2 Who should measure levels of distress and change?

2.2.1 Single-informant strategy

The second aspect to consider when measuring the outcome is who we ask to respond to the questionnaires. According to Johnsen and Torsteinsson (2012), the growth and development of CFT were partly as a reaction against the dominating medical model within mental health practices. The medical model implies that the therapist is the one to suggest or prescribe a treatment that will help the client, based on the therapist's exclusive knowledge. The medical model distributes power and authority to the expert and places the client, more or less, as a passive recipient. Further, based on the medical model, the one to report the levels of distress and possible developments in therapy is the therapist as he or she is the only one with knowledge of the disorders and expected treatment outcomes. CFT developed to a great extent as a reaction to this medical model, trying to distribute power and authority more evenly between clients and therapists. As a result, clients and therapists should collaborate when doing assessments and making decisions in and for treatment. This collaborative approach was developed with humility and respect for the client's own goals and means (Tilden, 2017). Tilden (2017) writes that embedded in this concept is the therapists' strong belief and trust in the client's wisdom and resources. Within this collaborative approach, the therapist relies on the client as the expert of herself, her problems and change, a perspective also shared by Duncan, Miller, and Sparks (2004), Ulvestad (2007) among others. Within CFT, the collaborative approach is the leading perspective (Lorås & Ness, 2019), and many therapists would actively value the perspective of the client higher than their own perspective of a problem or a change. The question was how to give more power to the voice of the client. One logical solution was to have the clients report levels of distress and possible developments in therapy and then give this information to the therapists, as has been done in this study. However, it is not the only possible option within CFT.

2.2.2 Benefits and challenges of the multi-informant strategy

Self-report outcome measures are very common, but some argue that the results are more compelling if an outside evaluator has been involved (Comer &

Kendall, 2013). One reason for this is that it is possible to argue that even if the client is the one closest to their life, their answers could be biased. Patients could possibly evaluate their improvement as more significant than an outside observer would notice (Ogles, 2013). *Multi-informant strategy* (Comer & Kendall, 2013), where data is collected from multiple reporters (e.g., client, family members, peers, teachers, observers, therapists), is, therefore, a suggested solution. The benefits of using a multi-informant strategy are that we get more reliable data (if a client and an outside observer identify the same change) and more nuances (when the answers from a client and an outside observer differ). The latter leads to a concern with a multi-informant assessment that discrepancies among informants are expected (Comer & Kendall, 2004) and is an issue researchers have to handle properly. One issue of concern is that also assessments of outside observers can be biased. Indeed, Walfish, McAlister, O'donnell, and Lambert (2012) found that therapists are overly optimistic about the improvements of their clients. Solem (2002) suggests using the multi-informant strategy but emphasizes that outside observers will need training to accomplish a high level of interrater reliability.

2.3 When to measure change?

Most research within psychotherapy has been on the outcome and not on the process (for reviews see Ogles, 2013 and Sexton et al., 2013). Outcome research refers to investigating the differences in scores from pre- to post-treatment measurements. Process research, on the other hand, is investigating the content of the treatment components associated with change that occurs in the therapeutic process from pre to post-treatment. Process-outcome research seeks knowledge on the process that is associated with the outcome by investigating predictors/moderators and mediators/mechanisms of change. The need for process-outcome research is emphasized by the fact that even though psychotherapy is effective for a large portion of those people seeking help (e.g., Lambert, 2007; Sexton et al., 2013), about half of the clients do not benefit from therapy, based on the measures used. Some even get worse (Ogles, 2013). To investigate the processes and measures that are completed frequently by the clients during treatment are needed. In an attempt to better understand, and explain, why some clients do not benefit from therapy while others do, an empirical approach that delves into the dynamics behind the psychotherapeutic

process has evolved over the last twenty years, named Routine Outcome Monitoring systems (ROM); (Lambert, 2010; Ogles, 2013; Tilden & Wampold, 2017). It has also been named patient-focused research (Lambert, Hansen, & Finch, 2001) due to its focus on the patient's responses and how this can inform the therapist in real-time therapy. Pinsof (2017), Lambert (2001), Duncan (2003), and Miller (2003) all suggest measuring at every therapy session. Schiepek and colleagues (2003; 2014) suggest daily measurements of clients' level of distress, as they emphasize the need for assessing the patterns of change. Regular measurements might be useful for assessment but burdensome for the clients, at least if they have to answer to comprehensive measures daily. For that reason alone, it is questionable if daily measurements are feasible in outpatient practice. Measuring at every session has, therefore, become the usual approach when investigating the process in therapy (Tilden & Wampold, 2017)

2.3.1 Routine outcome monitoring of clients' subjective experiences

ROM gives value to the clients' view of their life and their view of the change that might happen during therapy (Tilden & Wampold, 2017). Via standardized questionnaires that the client completes frequently during treatment. These data are fed back to the therapist, supervisors, and the clients themselves (Tilden & Wampold, 2017). Some ROMs even model the client data with his or her expected change trajectory based on data from similar clients with similar problems and levels of distress (Duncan et al., 2004; Lambert & Finch, 1999). Because ROM intends to monitor the outcome of the psychotherapeutic process as it evolves, it has provided valuable insights and help for both the client and the therapist (Duncan et al., 2004; Olkowska et al., 2018; Valla, 2014; Zahl-Olsen & Oanes, 2017). Moreover, some studies suggest that the use of ROM also improves the effect of the treatment (Anker et al., 2009; Reese, Toland, Slone, & Norsworthy, 2010; Tilden & Wampold, 2017) and that they identify in real-time cases that do not develop as expected (Lambert & Shimokawa, 2011).

However, a recent meta-analysis has cast doubt on this conclusion (Kendrick et al., 2016). These authors could not find evidence of a difference in outcome in terms of symptoms, between clients who used ROM in their treatment and clients who did not. Important to notice is that Kendrick et al. (2016) grade their comparison as of low quality because all the included studies were considered at high risk of bias. The risk of bias was connected to inadequate blinding of

assessors and significant attrition at follow- up. One possible reason for the discrepancy between studies that found good effects of ROM and studies that did not might be due to the level of simplification of the measures (Tasma et al., 2016; Zahl-Olsen & Oanes, 2017). In Paper 1, we tested what level of simplification that could be applied to a comprehensive ROM for it still to be clinically useful.

Several ROMs are available, and some of them are listed here: The Outcome Questionnaire (OQ-45; Lambert et al., 1996), the Systemic Clinical Outcome and Routine Evaluation (SCORE 15; Carr & Stratton, 2017; Stratton et al., 2014), The Treatment Outcome Package (TOP; Kraus, Seligman, & Jordan, 2005), The Partners for Change Outcome Management System (PCOMS; Duncan et al., 2003; Miller et al., 2003) and the Systemic Therapy Inventory of Change (STIC; Pinsof et al., 2009; Pinsof et al., 2015b). Since the ROM measure chosen for this study is the STIC, I will describe this measurement in more detail in the following.

2.3.2 Systemic Therapy Inventory of Change (STIC)

I will start by presenting the theoretical background for the STIC and why and how it was developed. Then I will present the scales and how they are intended to be used.

The theoretical background for the STIC is multi-systemic (Pinsof, 2017). This fact means that it captures the change in the primary systemic domains of a person's life: Individual, couple, family, and children. This approach allows for the investigation of how changes in one domain relate to change in others. For example, we investigated in Paper 3 how the change in individual and relational measures relates to change in work functioning. According to Pinsof (2017), the STIC prioritizes the client's experience and is client-focused, as opposed to therapist-focused. This choice is expressed by being a client self-report system (Pinsof et al., 2009) as opposed to the observational system. Indeed, Pinsof et al. (2009) state that if there is an outcome "bottom line," it is client self-report, implying that the client's view of the problems and the change that is in focus and the one given the strongest value. The STIC also has an integrative perspective that is consistent with the principles of family systems' thinking and practice (Pinsof, 2017). This perspective means that there is a focus on behavior,

cognition, and emotion, as well as a historical and here-and-now perspective of a client system. Pinsof and colleagues (2009) argue that this makes the STIC relevant to therapists from different intellectual traditions (e.g., behavioral, cognitive, affective, psychodynamic). This fact was also reflected by the research team, who developed the STIC who had members from different intellectual traditions (integrative, cognitive-behavioral, family systems, psychodynamic, and experiential) (Pinsof et al., 2009). The items were suggested by several clinicians, with each at least 20 years of clinical experience. The founder of the STIC is in his clinical practice working within the Integrative Problem-Centered Metaframeworks Approach (Pinsof et al., 2015a), but he claims that this theory does not have to be used by the therapist using STIC (Pinsof et al., 2009). However, the STIC "is designed to give therapists relevant and usable information that can influence their clinical decision-making and be shared directly with clients." (Pinsof, 2017, p. 87). Indeed, it is not the therapist who is the expert in understanding the answers presented by the STIC, but it is something the therapist and the clients should explore and try to understand together (Pinsof, 2017). STIC does not imply that there is a linear connection between what is happening in treatment and the clients' lives. The inventor of the STIC clearly states that it is not covering all areas of life, and that other aspects of peoples' lives might be even more important than what is captured by the STIC as he writes: "STIC clinical profile should never be thought of as a complete or definitive clinical picture of the client of the case and must always be supplemented and informed by the clients' personal narratives." (Pinsof, 2017, p. 92).

The STIC system consists of two client self-report questionnaires: The *Initial STIC* and the *Intersession STIC.* The Initial STIC is filled out by clients before treatment, which takes approx. 45 minutes to complete. It includes detailed demographic questions and six scales focusing on the different familial systems that comprise an adult's intimate life context. The Individual Problems and Strength scale (IPS) measures symptoms and wellbeing on eight factors. The Family of Origin scale taps into six factors of clients' view of the family where they grew up. The Relationship with Partner scale assesses the couple functioning on seven factors. The Family/Household scale assesses their experience of their current family on eight factors. The Child Problems and Strength scale delineates a parent's perception of a child's functioning on seven

factors. Lastly, the Relationship with Child scale asks the parent about his/her relationship with the child on three factors. Clients only fill out the scales that apply to their demographics, implying that a client in a current relationship with children living at home is answering several more scales than an adolescent still living in his or her family of origin. In this way, the STIC system adjusts to the clients' life situation, and at the same time, withholds the systemic perspective of the interaction between the individual and his or her close relationships (Pinsof, 2017).

The *Intersession STIC*, a briefer version of the Initial STIC that takes approximately seven minutes to complete, is filled out by clients before every subsequent session. All scales from the Initial STIC, except the family of origin, are used in the Intersession STIC.

Additionally, the clients evaluate their response to treatment, including progression, and also their alliance to the therapist and the treatment (Pinsof et al., 2009; Tilden et al., 2010). Client evaluations are processed into a report that is fed back to the therapists who can use this information together with the client as the basis for understanding and creating hypotheses in the clinical assessment of the current client. When the STIC data is displayed to the therapist and the clients on the internet-based platform, not all the details are presented at once. First, the total scales and *the Big Six*, i.e. the most clinical subscales, are presented (Pinsof, 2017). From that view, the therapist and client can jointly dig down via subscales, factors until reaching the bottom line layer that is the actual questions and the clients' answers. The feasibility of the STIC System procedure has been tested at The Family Institute, USA, and in two multi-site mental health consortia in Norway and Chicago (Tilden, Håland, Hunnes, Fossli, & Oanes, 2015).

In this study, STIC measures are used in Paper 1 and Paper 2. However, only answers from the Initial STIC were included. In Paper 3, we did not apply any measures from the STIC but choose only the sample of clients who used the STIC intersession during their therapy. The reason for not using any STIC measures in Paper 3 is that we intended to investigate the change in the STIC intersession measures to predict change on measures used in Paper 3 in a future study.

2.3.3 Level of simplification in measures

When complex human life is to be assessed, especially when assessed on groups of clients in quantitative research designs, we have to use standardized and simplified measurements. As explained, the aim of all these ROM systems is to help therapists understand the clients and their change during therapy. However, as mentioned, all attempts to standardize experienced change are reductionistic in its nature. The question is, therefore, how much reduction can be conducted on a ROM before the systems no longer help the therapist to understand the clients and the therapeutic process. Different levels of simplification of complex systems as a human being are presented in figure 1.

Figure 1: Different levels of simplification

A complex model of a man.

A simplified model of a man.

A very simplified model of a man.

What level of simplification is to be used was the main question assessed in Paper 1. Behind this question, there is a concern that simplified versions contribute to confuse and have therapists to misunderstand their clients. As Zahl-Olsen and Oanes (2017) point out, clients claim that despite the many questions asked, those questions do not always cover the areas of most importance to them. Furthermore, sometimes clients report that life has become better in some areas and worse at others, and if these are averaged and combined into one scale, these variations are leveled out and become invisible. Thus, as suggested by Zinbarg and colleagues (2018), a diverse measurement scale should allow for a better representation of the clients' opinions than a simplified one. As illustrated in figure 1, humans can be modeled at different levels of simplification. The model at the very right can be used if we want to count men, but it is not very suitable if we want to know anything more about these men. The illustration does, of course, not picture psychotherapy but exemplifies that any measure is a simplification and that with a high level of simplification, the details are lost.

Such details might be of the uttermost importance both for research and clinical practice.

2.4 How to measure change?

After considering what to measure, who to measure, and at what level of simplification we can apply to a measure, there are still several issues to address before the change on these measures is possible to assess. The main question I will address now is how to measure change. There are some questions that logically follow: What is change? Who should assess the change? How reliable is the change we have measured? And how big is the change measured in this study compared to what is found in other studies? Even if the main focus of this thesis is not on methodological aspects, I have chosen to include a discussion on several aspects. The reason is that I have used both cross-sectional and time-series data from a ROM system (the STIC), and I find the aspects of whom, what, when who and why to be related to this thesis. However, regarding the ontological and epistemological discussion mentioned earlier, discussing change using a predefined measure will contradict the convictions of some therapists and researchers within this field.

2.4.1 Change defined by whom?

An important issue when it comes to change is who is the one to define the expected or wanted areas of change as a result of a therapeutic process. From a medical perspective, the expected change implies that the problem is to be resolved or the mental disorder to be healed. From the client's perspective, this might not be the goal at all. For a client, it might be more important to increase relationship satisfaction than to decrease depression. Or it might not even be the goal to increase the relationship satisfaction either, but rather to leave the relationship. This variety can be handled in the single therapy but is very hard to capture when researching at a group level. When analyzing groups, researchers most often decide to measure all the clients using the same measures. This method is, of course, not fair to all clients and their individually defined goals for the therapy, or for that matter, for all processes of therapy that are included. However, if all clients are to define their goals and the direction they prefer them to change, it will be almost impossible to gain information on a group level. Standardized measures, like the ones used in this study, are reductionistic, and by

applying them to all the clients in a sample, no matter their reason for seeking therapy, is even further reductionistic. This situation applies specifically when assessing treatment that is not manualized, as in this study. Nevertheless, when implementing standardized measures on a group of clients, as performed in this study, it lowers the chances for finding significant results and also lowers the effect sizes (Field, 2018). It is an important limitation, but it is also an ethical aspect to this. Is it right to measure change by the use of a measure that people do not self-define as the primary outcome of the therapy? As mentioned by Zahl-Olsen and Oanes (2017), some clients might question that, even if most clients probably would agree that we use the information they have given to find out as much as possible to help themselves and other future clients. By doing so, we can analyze how different aspects of clients' lives are related. Paper 1 investigated what level of simplification of a measure that could be useful for research and clinical practice. Paper 2 investigated physical violence among couples and others in their family to inform therapists. Paper 3 assessed change on three different levels of the clients' lives and investigated how these were related.

2.4.2 Different measuring levels of change

Change is measured in different treatment contexts, and three different constructs are commonly being used: efficacy, effectiveness, and efficiency. *Efficacy* is the extent to which an intervention produces a beneficial result under laboratory-like conditions, whereas many as possible of confounding factors are controlled for (Wampold & Imel, 2015). There are two levels of efficacy. The *absolute efficacy* answers to the question: Does treatment work better than no treatment, while the *relative efficacy* answers to the question: Does this treatment work better than another treatment (Wampold & Imel, 2015). Efficacy studies have strong restrictions. First, the therapists are controlled to make sure they have the proper training to provide the treatment according to criteria for competence and adherence to protocol. Second, therapy is controlled to make sure that all clients have received the described treatment. Third, the clients are controlled to have only the disorder they intend to measure the effect of the treatment on, without comorbidities. As Wampold and Imel (2015) write, most often many of the ordinary clients are omitted from efficacy studies because they have comorbid diagnoses or difficulties. *Effectiveness*, on the other hand, refers to the benefits of psychotherapy that occur in community (naturalistic) settings and answers the question, "How effective is a treatment administered to clients in the 'real

world'?" (Wampold & Imel, 2015, p. 97). For clinicians, results from effectiveness studies are of most importance due to its closeness to real-life therapy. *Efficiency* is a related concept, with two different meanings in the literature. The first relates to the health economy (Bower & Gilbody, 2005) and answers the question, how much benefit at what costs? The second meaning is related to the individual client (Gullestad, 2001) and answers to the question, is this treatment efficient for this client?

A treatment that has great efficacy is not necessarily proven to have good effectiveness (Lambert, 2013b). However, this delusion has imparted the field and, particularly, on the ever-growing list of evidence-based treatments (Andrews, 2000).

2.4.3 What is good evidence of treatment effect?

Evidence-based Treatment (EBT) is a system of ranking treatments that has been abandoned by the American Psychological Association but is still being used and is a part of the discussions in the field of psychotherapy (see, e.g., Kazdin, 2008 and Sundet, 2019). I, therefore, find it important to present EBT before I present the modern evidence-based practice (EBP). The EBT system is founded in the American system of health care, Managed Care. One of the aims of Managed Care is to ensure quality in the therapeutic services and to ensure that the stakeholders pay for something effective (Ulvestad, 2007). It evolved from evidence-based medicine that had an enormous impact on medicine (Duncan & Reese, 2013). For a treatment to be counted as EBT, it must conform to three rigor requirements: (1) The approach has been shown to work using double-blind treatment and control groups with replication by at least two independent studies. (2) It has been translated into a treatment manual. (3) The treatment has been applied with specific client populations and problems, for example, depressed adolescents (Larner, 2004, p. 18). There are several levels of accreditation. To be top graded (level 1 = the treatment works very well), the evidence needs support from at least two large-scale RCTs, and these should not be conducted by the treatment developers due to the allegiance effect, as discussed in chapter 2.1. The requirements for evidence support are gradually reduced in the next four levels. The aim is to establish solid knowledge evidence supporting the recommendation of a particular method of treatment for a specific disorder.

EBT is a simplification of therapy that some researchers have called *oversimplification* (Clarkin & Levy, 2004, p. 214). Kazdin (2008) argues that EBT does not differentiate between outcomes, nor assess effect sizes, which implies that any significant change detected with an RCT count as much, no matter how big or small the effect size. Another aspect is that EBT favors treatments that are manualized and thereby easy to describe and measure (Johnsen & Torsteinsson, 2012). Treatments that are not as easily described in manuals and are harder to be tested in randomized controlled trials are discriminated by the system. This question has led researchers, practitioners, and health-care policy advocates into an ongoing debate on what counts as evidence and how they are to be used and integrated (Lambert, 2013b). Indeed, Duncan and Reese (2013) identify this debate as "perhaps the fiercest debate of our times" (p.489).

As a reaction to the negative effects of EBT, e.g., that the clients are easily being blamed for no treatment success and that only RCT studies are accepted, the American Psychological Association declared Evidence-Based Practice (EBP) to be "the integration of the best available research with clinical expertise in the context of client characteristics, culture, and preferences" (APA-Task-Force, 2006, p. 273).

EBP is a broader term than EBT, and in contrast to the EBT, EBP does not consider one source of knowledge as superior to another (Duncan & Reese, 2013). As Duncan and Reese (2013) point out, there is not one defined gold standard of research methodology. Effectiveness studies, process research, single-subject design, case studies, and qualitative research all bring important and valuable information when treatments are to be evaluated. Evidence within EBP is therefore considered from a much broader perspective than within EBT, holding that treatment outcome is related to a variety of contextual and relational factors, not just associated with offering the right method.

Further, EBP to a stronger degree than EBT targets the process-outcome research that seeks to learn more about the "why" and "what" of an effective treatment. APA's definition of EBP also includes the clinical expertise, meaning that therapists ultimately must use their clinical judgment to determine if they are to use an intervention that is based on research as effective for this particular client. This component of the definition highlights the inherent limitation of research findings – that knowledge of groups of clients does not have to fit one client in

that group. In particular, in APA's definition of EBP, it is emphasized that the client's knowledge about himself or herself (values, preferences, culture, the theory of change, etc.) must be included in the clinician's daily work. EBT's focus on the treatment intervention and not on who is delivering it or receiving it is perhaps the biggest difference between EBT and EBP.

This study is trying to answer questions within the paradigm of EBP. As such, and within that paradigm, it is important to assess what change is considered being a meaningful change. A statistically significant change in a group of clients does not mean that all clients in that group benefitted from the treatment. Neither does it mean that a client that did benefit would be able to notice the change in real life (Jacobson & Truax, 1991). Therefore, the change should be assessed for every client, one by one. How this could be performed is what I am going to discuss in the following chapter.

2.4.4 How to assess if a change is clinically significant?

Using quantitative questionnaires in psychotherapy implies comparing mean scores at the start and end of treatment for the sample (see, e.g., Langkaas, Wampold, & Hoffart, 2018). Since the analysis is performed on groups of clients, within-group and between-group variability are considered. The analysis produces a significance test. If the significance is greater than what could have resulted by chance, normally defined as $p < .05$, we call it significant. However, statistical significance does not provide evidence of clinical significance (Kazdin, 2008). Solely relying on statistical significance can lead to misunderstandings.

In comparison, clinical significance attempts to address if the treatment-related change was convincing and meaningful (Comer & Kendall, 2013). One procedure on calculating clinical significant change is developed by Jacobson and Truax (1991). This two-step analysis is performed on every client's data, one by one, and categorizes the clients into four outcome groups. The first step is to eliminate noise, and the second step implies using a clinical cut-off value. The clinical cut-off value is needed for separating clients considered being clinical and in need of treatment from those not clinical and hence not in need of further treatment. The mathematical formulas for calculating clinical significance and cut off-values are described in the method section in Paper 3 (See appendix, Paper 3 page 4 and 7). The four outcome groups are deteriorated, no-change, improved, and recovered.

The clinical significance is in the literature the most frequently used method for examining improvements (Ogles, Lunnen, & Bonesteel, 2001). However, this approach has its limitations, and the primary criticism has been that it does not account for regression towards the mean (Ogles, 2013). Regression towards the mean implies that repeated assessments with the same outcome measure, more extreme scores naturally become less extreme over time. Some researchers have suggested percent improvement instead (Hiller, Schindler, & Lambert, 2012), but since that method also has its disadvantages, it has not been widely used. Comer and Kendal (2013) ask for research using reliable change on what they call "real-world" referents (e.g., relationship satisfaction, role performance). We have followed this recommendation in Paper 3. In the analysis, both effect size calculations and clinical significant calculations were performed and compared for the clients' work functioning.

However, arguments have been made that clinical significance is not clinically informative (Kazdin, 2008; Langkaas et al., 2018). Langkaas, Wampold, and Hoffart (2018) argue that it cannot be applied to measures that have an upper and lower preferred limit, such as monitoring progress toward a target weight range in the treatment of an eating disorder. Secondly, they argue that a substantial proportion of clinical cases are prevented from ever achieving the status of recovered even if they do recover in treatment. The reason is that it requires that cases end up below the clinical cut-off (if higher is worse on the measure) to have the opportunity to achieve clinically significant change. The problem is that this cut-off does not perfectly discriminate between clinical and non-clinical cases. A study of a commonly used ROM, the OQ-45, suggests that about one in four clinical cases is prevented from ever being considered recovered because their scores are below the clinical cut-off even if they belong to the clinical sample (Bauer, Lambert, & Nielsen, 2004).

Finally, Langkaas, Wampold, and Hoffart (2018) argue that the concept of deterioration makes it sound that something has gone noticeably wrong – the client has become worsened through treatment. But, deterioration only indicates that there has been a reliable change away from the desired goal, and these authors argue that this "cannot indicate if the difference was of a magnitude of any practical clinical importance at all" (Langkaas et al., 2018, p. 248). They thereby suggest that clinical significance, which was intended to stop confusing

statistically detected differences with differences of clinical importance, does not manage to do so.

The solution presented by Langkaas, Wampold, and Hoffart (2018) is *induced difference,* indicating that observed progress is the likely consequence of the intervention. The point with this concept is that the observed difference must differ from the progress that would most possibly happen without treatment. For some conditions, we expect deterioration if no treatment is given, while we for other conditions expect improvement even without treatment. The induced difference would then be the change on a score that is different from this expected change. Langkaas, Wampold, and Hoffart (2018) argue that we need data on expected trajectories without treatment to compare with the results we get from ordinary treatment and incorporate that to ROMs so that clinicians get information of the induced change. However, no such system is available at the moment.

2.4.5 The possibility of arbitrary results when categorizing outcome

When applying boundaries on continuous measurement and defining some clients as responders and others as non-responders, we sometimes get arbitrary results (Senn, 2018). E.g., two clients experienced the same improvement, one passed the clinical cut-off value by one decimal point, and the other did not. Thus, the first client is defined as a responder and the other as a non-responder. In the Paper 3 study, we used four outcome categories, and the first client in the example above would be defined as recovered and the second as improved. However, arbitrary results might happen at the edges of the four categories as well.

Another aspect to consider on an individual level is the natural fluctuations in life. One client might have a bad day at the start of treatment and happened to have a good day at the end of treatment, defined as improved even if this change was arbitrary. As suggested by Senn (2018), frequent observations on the same client are needed to overcome this issue. This procedure is exactly what is done when using feedback instruments (i.e., ROM), where the clients respond on questionnaires frequently during a therapeutic process (see. e.g., Tilden & Wampold, 2017). The data from this study stems from feedback-informed treatment, and we have data from every session. However, it is outside the scope

of this study to analyze these frequent data, but we plan to do so in an upcoming study.

2.4.6 Considerations regarding sub-samples when measuring change in CFT

When conducting psychotherapy research there is always a question to find a suitable sample (see, e.g., Comer & Kendall, 2013; Ogles, 2013). Some aim for highly selected samples based on specific (and often narrow) inclusion criteria, with few if any comorbid diagnoses, yielding studies of high internal validity while other studies use samples from more ordinary practice, often defined as naturalistic studies. As mentioned, it is not evident that a treatment found to be efficacious within a clinical research setting will be efficacious outside this setting (Comer & Kendall, 2013).

Furthermore, when investigating CFT treatment, as I did in this study, there is another aspect to consider, especially when comparing the results with studies within individual therapy contexts. In the latter, due to priority guidelines in the public health system (e.g., Helsedirektoratet, 2015a, 2015b), only clients with high levels of distress for the disorder are offered treatment. For this reason, one may assume that their initial distress levels are in the clinical range, thus will qualify for inclusion to a clinical sample if they consent.

In CFT treatment, often both distressed and non-distressed clients are included in therapy and also in research because when couples come to therapy, it is often one in the couple who experiences the distress, while the other does not (Ness, 2017). Similarly, in family therapy, sometimes it is only one of the children who has a level of distress that leads the whole family into therapy (Lorås & Ness, 2019). Even if it is true that sometimes all clients involved have high levels of distress, I find it more probable in CFT compared to individual therapy that some CFT-clients have high levels of distress while others do not. CFT researchers thus need to make choices of whom to include in the analysis, and to the best of my knowledge, no unified conclusion on this dilemma has been agreed upon in the field (see. e.g., Langkaas et al., 2018; Sexton et al., 2013).

In a conversation I had with Bruce Wampold in 2018, he mentioned three different options when researching couple therapy. First, including all clients. Second, including only clients who initially indicated that they were in the clinical range. Third, including only the spouse in the couple who reported the highest level of distress, regardless of entering the clinical range or not. Such a

decision is of importance for all calculations of change. The first option, including all, implies that the change might be leveled out, and significant improvements are disguised. This situation might be the reason for the low to medium range effect sizes reported in many CFT studies (Sexton et al., 2013) compared to the high effect sizes often reported in studies on individual studies (Ogles, 2013). The second option, only including the clinical clients, implies priming the study by removing some of the clients who were present in the psychotherapy process that took place in the study. Accordingly, it is more probable to identify changes, and the effect size will get bigger than if all clients were included. Research on one of the ROM's commonly used for individual therapy, the OQ-45 (Lambert, 2015), have used the method of including only the initial clinical clients.

It is important to notice that the choice of only including the clients initially in the clinical range can lead to an inclusion of all clients from a family, only some from a family, or even none from a family. One problem with this method is that we do not know if several clients from the same family are included in the analysis, possibly confounding the results, especially if the number of participants is low. A few families with a very good outcome from the treatment could then increase the calculated effect size. Another issue is that some families might be excluded from the analysis because none of the clients indicated being in the clinical range at the start of therapy. This gap is discriminating some clients because of levels of distress on the measure used and some clients because the measure is not covering the areas of distress in their lives. Nevertheless, this type of selection is used by Lambert (2015) on individual therapy. In Paper 3, we chose to calculate the change using the total sample and the clinical sample and presented the results separately.

However, there is a distinct difference between cases included in research on individual therapy and CFT. In individual therapy, the clients are almost exclusively seeking help for their own problems, even if those problems could be viewed as caused by others or be in a relationship with other people. As mentioned, in CFT, the clients might seek help for individual problems, relational problems, or even problems with their children. Typical in CFT, all of those involved, can meet to therapy sessions, no matter if they themselves experience this phenomenon as a problem or not. It is, therefore, a higher possibility that one measure will not be able to identify the problems the clients

are seeking help for in CFT compared to individual therapy. In quantitative research, investigating change on groups of clients, one or a few measures are selected as outcome measures for all the clients in that sample (Ogles, 2013). This fact implies that some clients might be discriminated against based on the measures chosen.

The final option, suggested by Wampold, for couple therapy, was to include only the most distressed spouse. This method would include cases even if the measure did not indicate any of them being in the clinical range at the start of treatment, and thus makes sure that all cases are included in the analysis. However, this method excludes half of the sample, no matter if the person is highly distressed or not. We chose not to implement this method when assessing change in Paper 3, mainly because the sample size would have made it hard, if possible, to perform the types of analysis needed to reply to the research questions.

As described, the basis of CFT theory is systemic theory. Investigating change and, as just discussed, selecting samples solely based on individual levels of distress conflicts with the theory. Therefore, Paper 3 presents results both from the total sample and from the different samples of clients that were in the clinical range on the measures used in that study. In that way, it is made clear to the reader that the sample size is different when using the different measures, and that also the calculated effect size differs.

2.4.7 How to handle that a client's understanding of a concept might change from pre to post-treatment?

As mentioned, several aspects should be considered when investigating change, not only group change and the individual change. One aspect is Response Shift Theory (see, e.g., Schwartz, 2010, for a discussion) that takes changes within a clients' standard of a question into account. For example, a detected change from start to end of treatment might be a change of a clients' standard of the questions and not an objective change on these measures (see. e.g., Schwartz, Andresen, Nosek, Krahn, & Measurement, 2007).

Howard and Dailey (1979) point out that the response shift is a source of contamination of self-report measures. They argue that if pretest and posttest scores are to be comparable, a common metric must exist between the two sets of scores. In using self-report instruments, researchers assume that a client's standard for measurement of the dimension being assessed will not change from

pretest to posttest. If not, the measured change would reflect this shift, in addition to actual changes on the measure. Consequently, comparisons of pretest with posttest ratings could be confounded by this distortion of this internalized scale, yielding an invalid interpretation of the effectiveness of the intervention. This situation is especially interesting when one consequence of treatment is that the clients change their understanding of the variables we are measuring.

Howard and Dailey (1979), therefore, suggest to measure at pre and post as usual and add a question at post-treatment where the clients are asked to rate where they now, at post-treatment, would rate where they were at pre-treatment. They call this measure for "*then."* The difference between pre and then self-report ratings is referred to as response shift. Several studies comparing pre/post and then/post effects conclude that the standard pre/post design gives more conservative treatment effects (Howard & Dailey, 1979). It is thereby probable that the existent literature, relying on pre/post self-reported measures, underestimates the treatment effects. Treatment effects experienced by the client, and maybe even observable by the therapist, might not be detected when pre/post design is used. Studies have also found that then/post design is more in agreement with objective measures of change (Howard & Dailey, 1979). It is interesting that now, 40 years later, we still not normally include *then* measures in psychotherapy research.

One reason could be that the results from then/post-tests can be ambiguous (Schwartz et al., 2006). Based on the findings from the meta-analysis of Schwartz and colleagues (2006), several response shift methods have been developed (see. e.g., Oort, 2005; Schwartz, 2010; Schwartz et al., 2007) to overcome the interpretation difficulties. However, I was not aware of the issues of response shift at the time we started the data collection, and *then* questions at the end of treatment were, therefore, not incorporated in this study.

2.5 Decisions on what, who, when and how in this study

In this chapter (ch.2), I have explained and discussed several issues concerning measuring outcomes in CFT. I have discussed what, who, when, and finally, how to measure levels of distress and change. I will close this chapter by shortly state the decisions made in this study on those questions. This study started with collecting data from the clients themselves as they were the closest to report on

their own distress. However, I did not collect data from other sources, such as their employers or their therapists. Hence, a single informant self-assessment design was chosen. This study used a set of widely used standardized measures, and I did not develop any measures on my own. All these choices were made by the steering group of the main study before the start of my study (Tilden et al., 2019b). Among the available measures I chose measures with individual, relational and societal focus. Paper 1 assessed the individual scale of STIC to see if it was valid and reliable and at what level of simplification it would be wise to use in the search for answers to our research questions. Paper 2 used individual and relational measures to investigate domestic violence. Paper 3 used time-series data where individual and relational measures were used to predict work functioning. In the assessment of the change from pre- to post-treatment, the method of observed difference was used to assess change on group level and clinically significant change calculations to assess change on individual levels.

3 Domestic violence

So far, in this thesis, I have focused on the more general discussion on what, who, when, and how to measure. In this investigation of couple relationships within couple and family therapy, I will now move to the first of the two specific topics of this thesis, namely domestic violence.

3.1 Definitions of violence

Different definitions of violence exist, (see Ali, Dhingra, & McGarry, 2016) for a review). The World Health Organization (WHO; (2002) classifies violence into physical, sexual, and psychological abuse, and later, WHO (2015) includes violence against individuals, groups, and societies in the definition. However, the focus of this study is narrower, targeting domestic violence. Intimate Partner Violence (IPV) is a well-established concept (see, e.g., Antunes-Alves & de Stefano, 2014; Stith, McCollum, Amanor- Boadu, & Smith, 2012; World Health Organization, 2017), but it does not include the violence that exerts between others than the two adults in the family. Domestic violence (DV), also named domestic abuse or family violence, is violence by one person against another in a domestic setting (O'Leary & Woodin, 2009). It includes IPV, but is wider and includes violence also involving others than the two adults in the relationship. Both DV and IPV can take place in heterosexual or same-sex relationships, or between former spouses or partners (Stith, McCollum, & Rosen, 2011). DV can also involve violence against children, parents, or the elderly. It takes many forms, including psychological (verbal and emotional), physical, economic, religious, reproductive, and sexual abuse (O'Leary & Woodin, 2009). The violence can range from subtle, coercive forms to marital rape and to violent physical abuse such as choking, beating, female genital mutilation, and acid throwing that results in disfigurement or death (O'Leary & Woodin, 2009). The physical consequences of domestic violence range from minor scratches to invalidation or death. However, the psychological consequences are often described as the most damaging because the effects suffered are subtle and affect the person's identity and social support network (Entilli & Cipolletta, 2017).

The Norwegian government's action plan against DV defines this as follows: "Violence has many expressions and includes physical, mental, sexual and material abuse against a person to whom the perpetrator has a close relationship.

It may be a violation of the present or former boyfriend, cohabitant, or spouse. It may concern children, grandchildren, or other close relatives' abuse of the elderly; it may concern children who experience violence in the family and may be engaged in forced marriages and genital mutilation. In extreme terms, this violence takes life." (Author's interpretation; (Justis- og beredskapsdepartementet, 2014). This understanding involves the whole family, clearly demonstrating the serious consequences of violence and that violence in extreme consequence can lead to loss of life. However, this definition of DV is wider than the one used in this study in two aspects. First, this study focuses solely on the violence that is exerted within the current relationship and family. Second, this study focuses exclusively on physical violence. This focus does not imply that I consider the other types of violence as less important or influential in people's lives; in this study, it was mainly a question of narrowing the subject to be studied based on the available data. To investigate this matter, I chose to differentiate between violence exerted between the adult couple and the violence that included others in their family. Hence I use the terms *couple violence* and *family violence* to distinguish these from each other.

3.1.1 Types of domestic violence

DV has been a topic in research for many years, and Johnson and Ferraro (2000) made distinctions that have shown to be useful in understanding gender differences and prevalence in different samples, but also when considering treating couples experiencing partner violence. They defined four different types of couple violence. These are *common couple violence* (CCV), *intimate terrorism, violent resistance,* and *mutual violent control.* CCV is the type frequently found in community and clinical samples, while intimate terrorism and mutual violent control are more prevalent in samples from battered women's centers (Simpson, Doss, Wheeler, & Christensen, 2007). Intimate terrorism is where one partner, most often males, controls or terrorizes the other by the use of violence, while violent resistance is the definition of the response to this control, and thereby most often used by women. The mutual violent control is where both partners are intimate terrorists battering for control. CCV occurs between couples as a part of an argument, is evenly distributed among males and females, and Stith, McCollum, & Rosen (2011) argue that it is the most common type of violence in a clinical sample. CCV is distinctively different from the other three forms and connected to situations and not to the context of control and fear.

These four types can be simplified into two types, *situational* and *characterological* (Stith & McCollum, 2009). In situational violence, women are violent at least as often as men, while in characterological males dominate. In clinical settings, it is of most importance to distinguish between situational and characterological violence because it is only advised to treat couples with situational violence (Gottman & Gottman, 2015; Stith et al., 2011).

3.2 Theories of domestic violence

A well-known description of the dynamics of couples violence is Walker's (1979) cycle of violence model that consists of three phases. During the first, *tension-building phase*, the tension escalates between the couple. The second, *acute-battering phase*, is where violence is exerted. The violence could be either physical, psychological, or sexual, or a combination of these. After this phase comes the *honeymoon phase*, where the batterer may be genuinely repenting and attempts to make up for the violent behavior and promises never to hurt the partner again. The focus of Walker's model is to understand what is happening here and now, not what has happened in the past of those in that cycle of violence. Cano, Avery-Leaf, Cascardi, and O'Leary (1998) present a theory of DV that explains the exerted violence by two factors. Their first factor is about what has been experienced in the past and their second factor about what happened in the situation leading to the violent act. Others have built on their theory and added more details to concrete aspects of people's lives. One example is the integrated model of DeMaris and colleagues (2003) that has aspects from the background, the relationship stressors, conflict management, and conditioning factors. Even if such theories are useful, they lack a deeper theory of what happens in the individual who reacts by violence.

A useful theory in this respect is the theory of Gausel and Leach (2011) that explains how people react to moral failure. In intimate relationships, such as family relations and couple relationships, there are extensive opportunities for experiencing moral failure, e.g., forgetting important appointments, trying to hide taboo emotions or actions, failing to act as intended or expected, or even using violence. How a person appraises his or her moral failure is essential in the model of Gausel and Leach (2011). They differentiate between how moral failure affects a person's social image (other's evaluation of us) and self-evaluation of

self and specifically if it is appraised as a general or a specific defect. If it is appraised as a general defect that can lead to condemnation from others, it will lead to self-defensive acts, either expressed as withdrawal to protect oneself from further condemnation or action. One of these actions could be physical violence.

In the theory of Gausel and Leach, if the moral failure is appraised as a self-defect that is understood as mendable, the person will try to repair the moral failure. However, if moral failure is appraised as a self-defect that is global to the person's self-image (e.g., the thought that "I am bad"), it would lead to self-defensive acts, violence being one possibility. This conceptual theory is applied to reciprocal violence in Gausel, Leach, Mazziotta, and Feuchete (2018), a type of violence that is suggested to be the most prevalent type of domestic violence (Fusco & Fantuzzo, 2009; Stith et al., 2011). Gausel et al. (2018) explain and empirically test that people who experience themselves more as perpetrators than victims are more prone to seeking reconciliation than people experiencing themselves mostly as victims. I find this specifically useful because it can guide us not only to understand domestic violence but also on how to treat and prevent those acts. Hence, I chose to frame the part of this study focusing on domestic violence within this latter theoretical concept.

3.3 Consequences of domestic violence

Being exposed to couple violence is a risk factor for several mental disorders (Coker et al., 2002; Golding, 1999; NKPH, 2014). Even though couple violence primarily occurs among the adults, it affects and includes the children severely. Children living at home often witness the violence that happens and are profoundly affected (Appel & Holden, 1998; Kimball, 2016; Slep & O'leary, 2005; Øverlien & Holt, 2018). A Norwegian national survey (Haaland, Clausen, & Schei, 2005) found that 30% of the children witnessed the couple violence that had occurred, while an American study (Fusco & Fantuzzo, 2009) found that as many as 95% of the children had been exposed to violence exercised within the family. Of these children, 75% had an active role in trying to influence the situation by contacting a neighbor, calling the police, or protecting the victims of violence with their own bodies. A meta-analysis by Kitzmann and colleagues (2003) concludes that 63% of children exposed to violence developed internalizing (e.g., posttraumatic stress disorder) and externalizing (e.g.,

aggression) problems. Hence, this group of children could be expected to become clients in children and adolescence psychiatric clinics. Furthermore, children who experience violence in their childhood have higher risks of conducting violence themselves (Raundalen, 2009) in addition to being exposed to violence in their adult lives (Renner & Slack, 2006; Wolf & Foshee, 2003; Øverlien, 2012). Hence, there is a high risk that the problem of violence passes on from one generation to the next.

In the meta-analysis by Kitzmann and colleagues (2003), they conclude that being a witness to violence is as harmful to the child as being directly exposed as the object to violence. When a child has experienced domestic violence, as a victim or a witness one or more times, the home is no longer the safe base, it is supposed to be for healthy growth and development. On the contrary, the home becomes a dangerous place, and the one(s) who should provide support and safety becomes a threat. The violence between caregivers spreads fear, and when the child needs security the most, caregivers are not present for them. "In this way, the children become orphaned in a situation where they sorely need the caregiver" (Author's translation; (Middelborg & Samoilow, 2014, p. 21). Another factor is that some children have been informing adults about ongoing domestic violence, but experienced that no action was taken. Indeed, several civil cases have been raised towards the child protective services in Norway due to lack of reaction in cases where maltreatment was detected and reported (e.g., Sandvig, 2009).

As mentioned, to be exposed to violent assaults increases the risk of serious psychological and psychosocial problems (NKPH, 2014), and is a problem that affects many. Hence, it is nearby to assume that those seeking mental health care have experienced more DV than the average population. To the best of my knowledge, there is no overview of how many of those in need of assistance from the mental health services have such experiences, but there are reasons to assume that the proportion is significantly higher than what is reported (Ormhaug, Jensen, Hukkelberg, & Egeland, 2012).

3.4 Prevalence of couple violence in the population in general and in clinical samples

WHO refers to national population surveys from 48 countries, where the results varied widely by country, but that as much as 69 % of women, in some countries,

state to have been subjected to physical abuse by a partner at some time in their life (Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002). Typically for this field is that the focus is on women who experience violence from men (Stith et al., 2011) even if violence against men is found to have similar prevalence (Capaldi, Knoble, Shortt, & Kim, 2012; Haaland et al., 2005; Jose & O'Leary, 2009; Thoresen & Hjemdal, 2014). However, Amnesty International (2013) writes that, even though women carry out violence against men in close relationships, it is mainly men who expose women to serious violence in close relationships.

In this thesis the focus is on violence among clients seeking CFT and the prevalence in my sample might differ significantly (i.e. is assumingly higher) from the general population, especially since we know, as mentioned, that being exposed to violent assaults increases the risk of serious psychological and psychosocial problems (NKPH, 2014) and hence might lead them to seek psychotherapeutic help. Studies from the US have found that up to 61% of CFT clients seeking couple and family therapy have experienced couple violence (Cascardi, Langhinrichsen, & Vivian, 1992; Jouriles & O'Leary, 1985; O'Leary, Vivian, & Malone, 1992; Vivian & Malone, 1997). I have not been able to find studies of the prevalence of couple violence within clinical samples outside the US, but according to Ormhaug et al. (2012), I have reason to assume that the proportion is similarly significant also in other western countries, such as Norway. Indeed, a retrospective study of 18-year-olds who have been treated in the department of child and adolescent psychiatric services (BUP) in Norway shows that 60% of this group had experienced DV, but this was only registered in 0.4% of the cases in the BUP's medical report system (Reigstad, Jørgensen, & Wichstrøm, 2006). This finding highlights an important aspect of working with DV, namely how this can be uncovered during treatment. I will come back to that in chapter 3.6.

3.4.1 Prevalence of domestic violence in Norway

The Norwegian Directorate for child and family affairs (Bufdir) states that as many as one in five women and men in Norway has been victims of violence or sexual assault from their partner (Bufdir, 2017). In a recent nationwide survey of violence (Thoresen & Hjemdal, 2014) they found that about as many men as women (14-16%) reported less serious partner violence in their lifetime (pinched, scratched, pulled by the hair or slapped), but also, there are many children in

these relationships. However, more women (8%) than men (2%) had been subjected to serious violence from a partner (kicked, strangled, beaten up). What was defined as less serious partner violence in that study appeared to be common among men and women. On the other hand, there are far more women than men exposed to serious violence (Justis- og beredskapsdepartementet, 2014).

Reports from the woman shelters indicate that it is the ones who have children who wait for the longest before they seek help (Krisesentersekretariatet, 2017). This fact means that children may live for a long time as witnesses, and perhaps even as direct victims of violence in the families, before the issue is addressed. A Norwegian survey conducted among 7000 18-19-year-olds in upper secondary schools shows that 10% had witnessed violence against at least one parent (Mossige & Stefansen, 2007).

3.4.2 Change in the understanding of domestic violence

There have been changes in the understanding of violence over many centuries. Back in the 15th-century, violence was understood as when someone from the lower levels of the hierarchy used force against those higher in the hierarchy, while it was called chastise when it went the other way (Sandmo, 1999). Chastise was also the label for when a man used physical force on his wife or children (Pape & Hjemdal, 2004). Until 1868, it was legal in Norway for a man to y punish his wife physically. Just over a hundred years later (1972), the Penal Code, which gave parents the right to punish children, was abolished. Especially through the women's liberation era in the 70th and 80th, many books about violence against women were published (Middelborg & Samoilow, 2014), several cases were tested legally, and a change in attitude emerged. In 1978, the first women's shelters were established in Norway, called *crisis centers*. The founders of the early women's shelters presented violence as an extreme form of oppression of women (Middelborg & Samoilow, 2014), a picture of domestic violence that is still present. Founded in feminist theory, specialized treatment offices were established in 1987 (Alternativ til vold/ English: Alternative to violence). Parallel, this work was fueled by the fight for equality between the genders that also impacted the understanding of DV. Middelborg and Samoilow (2014) state that when violence is viewed in this perspective, there is a risk that other types of violence not fitting into this picture become invisible. For instance, violence against children did not get as much attention, and violence from

women towards men was considered nonexistent. However, national data, from as early as 2003, clearly indicate that female to male violence is as prevalent as the opposite way around (Haaland et al., 2005). This finding is also confirmed internationally (Desmarais, Reeves, Nicholls, Telford, & Fiebert, 2012a, 2012b). From 2010 onwards, there has been zero tolerance in Norway for all forms of physical punishment for educational and upbringing purposes. It is only recently that other forms of violence than physical violence have been a part of the debate of violence (Middelborg & Samoilow, 2014). The current understanding of violence includes a variety of types as described in the definitions of violence already mentioned.

Treatments for domestic violence in the US began in the late 1970s and were closely aligned with the battered women's movement (Rosenbaum & Kunkel, 2009). The early treatment programs were for males only and focused on stopping men's violence, protecting the female victims, resocializing men toward equality with women, and getting perpetrators to take responsibility for their behavior (Mederos, 2002). Stith and McCollum (2009) describe the change in understanding among developers of treatment programs. They describe that in the early years, the notion was that all offenders were males and that they used violence to gain power and control. Further, women were seen as victims that only used violence in self-defense, and if they stayed in the relationship it was because they had achieved "learned helplessness." At that time the only intervention was group therapy sessions for men. As the knowledge of the diversity among domestic violence grew (i.e., the mentioned differentiation between situational and characterological violence) and evidence that women also used violence, treatment interventions have developed. Even though 68 % of the 44 US states with standards for domestic violence interventions prohibit funding of any program that offers CFT as a primary mode of intervention, CFT treatments for domestic violence have been developed in the US (Stith & McCollum, 2009). A question of interest is, thus, how these CFT treatment programs are funded. I do not know the answer to this question but notice that the first book describing a CFT treatment program for domestic violence in Norwegian was published in 2014 (Middelborg & Samoilow, 2014) and that it was based on a treatment program that had been offered to the public for several years at that time.

3.5 Recommended treatments for domestic violence

Individual or group therapy has been and is the most common treatment offered both for the victims and the offender (Stith et al., 2011). Couple therapy has commonly not been recommended when violence is occurring, and it is suggested that it might even contraindicate (Stith et al., 2012; Stith et al., 2011) mainly of two reasons: Firstly it may increase the risk of escalating the violence (Stith et al., 2011), because one partner might disclose actions not wanted by a perpetrator to be disclosed, thus escalating a conflict possibly leading to more violence. Second, it may mask the violence (Middelborg & Samoilow, 2014) because the nature of conjoint sessions in CFT imply that the partners in the couple do not tell the truth about what happened as they are afraid of how the other partner might react. However, group therapy for perpetrators of DV has found to be associated with increased risk of violence as well because of male bonding in groups, where the group members reported an increased level of anger after the group sessions because they felt mandated to and belittled by the fact that they had to attend to group therapy (Stith & McCollum, 2009). The male-only treatments also have shown to have at least a 50 % drop out rate (O'Leary, 2001) and have little empirical support for being effective (Rosenbaum & Kunkel, 2009). Several studies have documented that CFT treatments are effective on couple violence (Fals-Stewart, Kashdan, O'Farrell, & Birchler, 2002; O'Farrell, Murphy, Stephan, Fals-Stewart, & Murphy, 2004; Schumm, O'Farrell, Murphy, & Fals-Stewart, 2009). Karakurt, Whiting, Van Esch, Bolen, and Calabrese (2016) conclude in their meta-analysis that couples' violence can be significantly reduced through couple therapy. A finding supported by the review of Carr (2019) that identified several different CFT approaches suitable for treating domestic violence (e.g., behavioral and solution-focused CFT). Indeed, Vall, Seikkula, Laitila, & Holma (2016) argues that CFT treatment makes it possible to work on the alliance between the spouses, help them become aware of escalation processes and to discuss strategies to handle possible future violence with couples. However, this study is not investigating CFT treatment tailored specifically for domestic violence but is investigating DV in a sample of clients seeking CFT treatment.

3.6 Uncovering violence

Middelborg and Samoilow (2014) write that it is important that couple therapists have strategies to detect whether the violence that occurs in the relationship as

they argue that clients do not easily inform their therapist about this issue when they seek help. Physical couple violence is a sensitive topic to investigate because it is socially and legally not acceptable (World Health Organization, 2015). Furthermore, Middelborg & Samoilow (2014) argue along with Stith, McCollum, and Rosen (2011) that some clients will not report couples' violence due to the risk of more violence if the violent partner does not appreciate that this information is shared in therapy. This theory is supported by the results of Ehrensaft and Vivian (1996), who found that fewer than 10% of couples experiencing couple violence spontaneously reported or identified the violence as a presenting problem when seeking therapy.

One could assume that the agencies reported to the police at least a reasonable amount of these cases where they uncovered DV, but this does not seem to be the case. In 2014, the family counseling services in Norway registered to work with 32.695 cases (SSB, 2016). In addition, a large number of CFT cases were also treated in other services, but only 3075 cases were reported to the police the same year. If the prevalence of DV in Norwegian CFT samples is as high as the 61% found in CFT in the US (Cascardi et al., 1992; Jouriles & O'Leary, 1985; O'Leary et al., 1992; Vivian & Malone, 1997) this indicates that most of the cases with issues of violence are not reported.

When public employees learn about family violence and/or suspicion of abuse and negligence, they are obliged to report this to the child protective service (Barnevernloven, 1993) and/or the police. Thus, the most obvious reason for the low number of reports to the police would be that the professional helper is not aware of the violence. Another explanation may be that therapists refuse to raise violence as a topic in therapy (Ehrensaft & Vivian, 1996; Middelborg & Samoilow, 2014; Stith et al., 2011). Such a reluctance could be inspired by the social constructionist idea commonly present among many CFT therapists for not theming anything but on what the client wants to focus. Thirdly, therapists might wait as long as possible to report these issues to establish a strong alliance and thus obtain the possibilities of working with them therapeutically. For the two latter interpretations, this may be related to therapists working in a post-modern tradition (Anderson, 1997), where it is essential to hold a neutral attitude, implying not taking a side in therapy. This key value is seen in the work of CFT, and revealing the violence can make it difficult to have a neutral attitude

(Middelborg & Samoilow, 2014). Hence, there exists a dilemma between the ethical obligation as professionals to reveal and report client violence and neglect while at the same time CFT professionals may share the theoretically based reservation on inviting such topics to the table. However, it is also possible that therapists "close their eyes" to DV because otherwise, it would imply a commitment to take actions that could destroy the therapeutic alliance and lead to much more work with that case. Husso et al. (2012) identified that professional practitioners are ambiguous about how to handle domestic violence. However, Virkki et al. (2015) found that the attitudes used to justify lack of intervention in domestic violence could change to become more positive. Furthermore, Husso et al. (2012) indicated that lack of knowledge about the causes and effects of domestic violence often leads to feelings of inadequacy and frustration among professionals. Paper 2 is partly answering this need for knowledge as it investigates predictors of physical violence in close relationships.

3.6.1 Considerations for standard assessment of violence at pre-treatment

One way to help therapists to be aware of DV is by the use of routine screening (viz. ROM), especially since some research indicates that persons are more likely to reveal family violence when responding to a questionnaire compared to when questioned face to face by a therapist, e.g., Andersen & Svensson (2013); Ehrensaft & Vivian (1996); Zahl-Olsen & Oanes (2017). Several screening instruments have been suggested as a standard assessment of violence at pre-treatment (Stith et al., 2011). To our knowledge, no standardized screening of DV has been widely implemented within CFT before the Systemic Therapy Inventory of Change (STIC) (Pinsof et al., 2009) was introduced at selected CFT sites (STIC is described below). Within STIC all clients specifically were asked about couple violence and family violence, and this was reported to the therapist. However, using ROM does not guarantee that violence is detected. First, because people might restrain from reporting it. Second, if this reporting is addressed by the client, there is no guarantee that the therapist reads the report. Third, it depends on how the therapist interprets the reported violence and thereby choose to use that information in therapy. If he/she chooses to use the information in the session, in line with the intention of the user involvement in ROM, this implies discussing with the clients their understanding of the basis for conjointly planning the treatment. However, most CFT therapists do not have such a tool

and knowledge. Hence it is a task to inform and educate CFT therapists of signs to be aware of (i.e., predictors) that could serve as a great help detecting couple and family violence. Furthermore, knowledge of predictors of couple and family violence can help therapists to understand the complexity of this type of violence and thereby possibly provide more appropriate therapy.

3.7 Predictors of couple and family violence

Predictor studies of couple violence have mainly been conducted on non-clinical samples, leaving only a few addressing clinical CFT samples (for a review see, e.g., O'Leary & Woodin, 2009). Unlike many other health problems, few social and demographic characteristics define risk groups for intimate partner violence (Jewkes, 2002). Nevertheless, couple violence relates to other personal and relational factors. Two such characteristics are low income and low level of education that have been found to be associated with a higher prevalence of couple violence (see. e.g. Pan, Neidig, & O'leary, 1994; Stith et al., 2011). Furthermore, it has been determined that physical couple violence is associated with alcohol and substance abuse, both at present and in adults' family of origin (see e.g., Coker, Smith, McKeown, & King, 2000; Fals-Stewart, 2003; Fals-Stewart, O'Farrell, Birchler, Córdova, & Kelley, 2005; Stith et al., 2012). Experiences of alcohol abuse in childhood have been found to be associated with a higher risk of problems in need of treatment in adolescent and adult life (Dube et al., 2001). In addition, the use of violence in close relations is confirmed to be closely related to anger (Simpson et al., 2007).

Even if a close relationship between physical and sexual abuse has been found in several studies (e.g., Simpson, Atkins, Gattis, & Christensen, 2008; Simpson et al., 2007; Stith et al., 2012) I could not find any study that investigated the relationship between sexual satisfaction and couple violence. Issues of sexual satisfaction are common in CFT treatment (Almås & Pirelli Benestad, 2017). Since good sexual satisfaction is found to be associated with relationship satisfaction, love, commitment, and stability (Butzer & Campbell, 2008; Sprecher, 2002; Young, Luquis, Denny, & Young, 1998) it is reasonable to believe that there is less couple violence when the sexual satisfaction is high. Most couples argue and fight over practical issues like household chores, and, commonly, some of these issues occur repeatedly (Gottman & Silver, 2015).

Gottman and Silver (2015) argue that the underlying issues are often values or expectations: e.g., if the expectation of one partner about the other partner's household chores is higher than what the other expects of his- or herself, there is a likelihood of conflict, leading to tiring arguments with high tension or even violent outcomes (Gottman & Silver, 2015). However, different people have different conflict management styles that have been assessed in relation to couple violence (see, e.g., Gottman & Silver, 2015; Simpson et al., 2007; Vivian & Malone, 1997), and the findings suggest some styles of conflict management (i.e., use of hostile, verbally aggressive and avoidant conflict styles) to be far more associated with violence than others, but there is a consensus that self-control is viewed as a positive asset in high conflict (see e.g., DeWall, Baumeister, Stillman, & Gailliot, 2007; Sellers, 1999). As mentioned, the knowledge of couple and family violence within a clinical CFT sample is limited. There are even fewer studies of predictors of couple and family violence within this group of clients, and to my knowledge, no studies where family violence has been included. Hence, this study contributes to the knowledge of the prevalence of physical violence in a CFT sample and might also contribute to the clinical work of therapists to uncover ongoing physical violence.

4. Work functioning

As mentioned in the introduction, few studies within CFT has focused on societal aspects (e.g., sick leave or work functioning) of the clients' lives. Work functioning is one aspect that is of great importance to society, and this is, in particular, associated with individual and relational aspects of people's lives. Work functioning has therefore been chosen as the second focus in this investigation of couple relationships among couples receiving CFT. Waddell and Burton (2006) argue that "There are economic, social and moral arguments that work is the most effective way to improve the well-being of individuals, their families, and their communities" (p. vii). In their extensive review of the effects of employment Waddell and Burton (2006) argue that there is a strong evidence base showing that work is generally good for physical and mental health and well-being. They further write that "worklessness is associated with poorer physical and mental health and well-being" (p. ix). Waddell and Burton (2006) also suggest that work can be therapeutic and can reverse the adverse health effects of unemployment. According to Waddell and Burton (2006), this is true for healthy people of working age, for many disabled people, for most people with common health problems, and social security beneficiaries.

The social welfare system in most Western countries implies the presence of an economic safety net from public insurance for unemployed people, those people on sick leave, disabled, or retired. These public insurance welfare costs have, however, reached a level that concerns bureaucrats and politicians in most countries of the OECD (Organization for Economic Co-operation and Development) (Setzer & Rürup, 2013). For this reason, efforts are made to understand and establish means counteracting this development from political as well as from professional perspectives. For instance, recent research performed in Scandinavia found that more than 50% of sick leave was due to social problems (e.g., family-related distress or mental health issues) rather than to medical issues (Aronsson et al., 2015). The portion of those on sick leave due to mental health issues has been steadily increasing (Hensing et al., 2006) throughout the OECD countries (OECD, 2003). For instance, psychiatric disorders were the most common reason for long-term sick leave (Henderson, Harvey, Øverland, Mykletun, & Hotopf, 2011).

The research reviewed by Waddell and Burton (2006) identified that employment is generally the most important means of obtaining adequate economic resources, which are essential for material well-being and full participation in today's society. Work also meets important psychosocial needs in societies where employment is the norm, and work is central to individual identity, social roles, and social status. Waddell and Burton (2006) also found that there is a strong association between worklessness and poor health. This phenomenon may be partly a health selection effect, but they claim it is also, to no small extent, a cause and effect relationship. There is strong evidence that unemployment is generally harmful to health, including higher mortality, poorer general health, poorer mental health, and higher medical consultation (Waddell & Burton, 2006). It is established knowledge that occupational functioning is associated with mental health and the relationship with one's family. For instance, the individual consequences of dropping out of work may be reduced personal wellbeing, including poorer finances (Warr, 2003), depression (Adler et al., 2006; Lerner & Henke, 2008), relationship distress (Novak, Sandberg, & Davis, 2017) and reduced family functioning (Rotunda, Scherer, Imm, & Deleon, 1995).

On the other hand, the experience of relational distress has been found to reduce a person's school and work functioning (Whisman & Uebelacker, 2006). Research has shown that suffering from such as depression or anxiety (Bilsker, Gilbert, Myette, & Stewart-Patterson, 2005) and relational discord (Aronsson et al., 2015) reduce people's work functioning.

In sum, the connection to what is often the focus in CFT treatment and work function is thereby close. CFT has for instance been found to be effective for treating symptoms associated with lower work functioning, such as depression and anxiety (Carr, 2014a; Crane & Payne, 2011; Gurman, 2008, 2015), as well as dyadic and family problems (Sexton et al., 2013). Hence, one may assume CFT to be a suitable method in the effort to increase clients' levels of functioning at school and work through improving their relationships and reducing their depressive and anxiety symptoms. A reasonable theory is that the high conflicts leading parents in a family to seek therapy impacts on diminished function at work. After the relational conflict has been resolved they can probably contribute as normal again at work. As it is common to have a focus on CFT for both individual and relational aspects of people's lives, it is reasonable to theorize that

both aspects interact. Even when not being a therapeutic focus and thus not considered (and measured) as an outcome in CFT, it is likely to assume that work functioning is correlated with the change during that therapeutic process. Despite this, I have not been able to find any study that explicitly examines work functioning within the field of CFT. Paper 3 investigates work functioning from pre- to post-treatment and is in this study the only one using longitudinal data.

5 Method

5.1 Me as researcher, therapist and local administrator

Within the field of qualitative research, the researcher is not seen as someone who is outside the object he or she is investigating (McLeod, 2013; Punch, 2014). In this tradition, it is important to present what theoretical professional and personal background that might influence what is investigated. In quantitative research, such a focus on the researcher is not common. Some would argue that the researcher is not of interest since he or she is objectively trying to investigate something. However, Comer and Kendall (2013) state that the researcher's expertise has an important role in the production of the results. Indeed, I was one of the therapists participating both in the Pilot study and in the RCT study using STIC in my clinical practice. In the RCT study, I became the local administrator for the research project and part of the Norwegian STIC steering group. Additionally, I was the supervisor for the therapists in the department for child and adolescent mental health, Sørlandet Hospital, in the use of the feedback instrument (STIC).

Wampold and Imel (2015) point out that results from research on a specific method that was conducted by the developers of this method could be biased. They conclude that results from effect studies performed by method developers generally present higher effects compared to research on this method conducted by independent researchers who did not invest in the development of the method. The relevance from this research to the assessment of my role as a researcher in this project is that even if I was not the developer of the STIC system, I personally met him and his colleagues several times. Hence, I was not completely independent of the developer. Indeed, my colleagues and I suggested improvements to the online system that was later incorporated. Thus, on a continuum from no connection to the method, to being the one developing it I am possibly in the middle. However, my involvement made the system that was not available to the general public, available to me, and thus also the data. Hence, without my connection to the developer and the development, I would not have been able to do this study.

Another impact of me as a researcher in this study is that my role for some years has been as a CFT therapist. Thus I am interested in CFT and choose to investigate clients in CFT treatment instead of, for example, individual treatment.

The choice of methods for analysis is also colored by the knowledge that my supervisors and I had at the time they were conducted. New methods have been learned during the process of this study, and if this study had started today most possibly other choices would have been made. Most assuredly, I would have been using Bayesian methods of analysis if I had started today and not three years ago. One of the differences would be that I could have presented the probabilities for our hypotheses are actually based on our data, rather than probabilities of the data given the hypotheses. One example is that in Paper 3 the results could have been that the probability of correctly estimating the work functioning of a client is 60% when only using the individual measure, but at high as 75% when using individual and relational measures (the numbers serves just as an example and are not calculated). Choices I made in the process of this study were based on prior research, my own and my supervisors' personal preferences, and perhaps professional agenda.

Another aspect that has inspired me to conduct this study is that I am working in a field where different approaches are competing for ground. CFT departments have been closed in several locations in Norway during the last few years, giving space to individual treatments with stronger empirical support. I evaluate this change as tragically and understand this as a call for more CFT research, especially quantitative research, that this study is an example of.

I have been conducting research on data collected partly in the department in which I am working. My role was a local administrator of the RCT study and the supervisor in the use of STIC as a clinical tool. Further, some of the data is collected from clients participating in therapy where I was the responsible therapist. All these aspects need a discussion. Due to my different roles in this study, some of my colleagues could have felt pressure to participate in the study; however, no one was forced, paid or obliged to participate. My role as the local administrator, supervisor and in the end, also the one who entered the research role might have had an impact in different directions. The decision to give me the role as the researcher was made late in the data collection process. Some colleagues probably viewed this as positive because they knew me and trusted my handling of data from their clients. It might be scary to supply a researcher at a different site with data, knowing he or she will scrutinize the data from their therapies. On the other hand, it might have been a concern to have a colleague to investigate their therapy as well.

Being physically close to the data collection at ABUP made it easy for me to follow up on cases and questions which the therapists and clients had during the data collection. At some stages, we had quite a lot of technical problems, e.g., not being able to log in via electronic platforms to the website for completing the questionnaires and getting the reports displayed, and therefore daily hands-on from me was crucial. I had regular e-mail contact with the staff at the Family Institute in Chicago and participated in online problem-solving meetings not only for cases connected to ABUP but for all participating agencies. Locally I came to the therapist's offices to fix or explain, but also call or meet clients who had difficulties submitting their answers or had questions about the system or the research. Being so close to the clients and therapists gave me an understanding of the data that helped me to prepare the dataset for analysis. This knowledge was also especially helpful when I was to analyze missing data.

5.2 Participants

Data for this study stem from two separate, but linked studies, one pilot study (Tilden et al., 2015) and one RCT study (Tilden et al., 2019b). The RCT study collaborated with a parallel and similar multicentre study in Evanston/Chicago, USA, enabling multisite and multinational comparisons. The initial plan was to collect follow up data at 6 and 12 months, but because of a low response rate, this plan was abandoned during data collection.

This study included data from four different sites representing the stepped level of care of CFT in Norway. The first level was represented by two low thresholds, outpatient clinics where no referral was needed (one in the pilot study and another one in the RCT study). The second level was an outpatient clinic organized as part of child and adolescence psychiatry, where referral from a general practitioner or the child protective services is needed. The third and final level is a highly specialized and national inpatient unit organized as part of adult psychiatry that offers clients intense family treatment for up to three months. Because my co-authors and I consider these three levels to be representative for how the public CFT services are organized in Norway, I believe that including this diversity of sites strengthens the ecological validity of this thesis and thus facilitates the generalizability of findings.

The Pilot study included 502 clients and 40 therapists (licensed psychologists, psychiatrists, clinical social workers, psychiatric nurses, etc.). Data collection

started in March 2010 and ended in April 2014. The data collection for the RCT study started in August 2013 and was ended in September 2016, including a total of 339 clients and 23 therapists. All therapists in the RCT study indicated that they identified them with family/systems theory "a lot" or "completely," except for one who indicated "a little" (therapist identification was not identified in the Pilot study).

The main aim of the RCT study was to investigate if there was any difference in outcome when the STIC was being used in therapy compared to treatment as usual (without the use of STIC). For this reason, the participants were randomized into two treatment conditions, (a) treatment as usual (TAU) and (b) TAU + the use of STIC. In the following I will use the terms *participants* and *clients* interchangeably. The randomization procedure was performed automatically by the Psychotherapychange.org website after the clients had filled out the pre-therapy battery of assessment. This procedure means that the first client in a given modality (individual, couple, or family) was randomized into either TAU or TAU + the use of STIC. The next client who was enrolled in that same modality was given the other condition. For every new client, block randomization was repeated, but both clients in a couple and all clients in a family were handled in the same situations. The clients in both groups of the RCT study completed the outcome package (STIC plus seven other questionnaires. See Tilden et al., 2019b for a detailed description) through the same system as the STIC questionnaires. All participating therapists were collecting data for both treatment conditions. This procedure was done to avoid therapist effects, meaning that because some therapists get better results than others and if by chance the better therapists were placed in one of the conditions this would impact the results systematically (Wampold & Imel, 2015). Further, no defined treatment manual was used other than training followed by certification in the use of the STIC ("STIC guidelines") when meeting clients in the STIC condition.

All therapists received regular supervision in the clinical use of the system. However, as a naturalistic study, achieved competence and adherence to protocol were factors for which there was no control. There were no restrictions on what treatment interventions to offer, or what to focus on in the therapy. The therapists were told to continue offering treatments as they usually did. The only difference was that when the therapists got a case randomized to use the STIC they had to

use feedback from the STIC system actively in the therapy sessions. Furthermore, to plan for generalization of the results, the clients were treated by a variety of therapists within ordinary clinical practice.

The clients participating in this study represent a diverse group of people seeking couple and family therapy. It includes adult individuals (in individual therapy), spouses (in couple therapy), single or couple parents, and adolescents from the age of 12 (in family therapy). Even if they were seeking help for different problems, they all had in common that aspects of their close relationships were perceived and viewed as severe.

In Paper 1, we used cross-sectional data at the start of therapy from the total sample (N=841), which is the entire sample from the Pilot Study and the RCT Study together. Similarly, we used this total sample in Paper 2; however, by excluding the 11 clients below the age of 18 (N=830). In Paper 3, we used longitudinal data from only one of the conditions in the RCT study, the feedback condition (N=165). Since the pilot study did not include any outcome package and lacked a clear definition of when treatment was ended, we could not include data from that study in Paper 3. The decision of only including data from the feedback condition in the RCT in Paper 3, was because only data from that condition have data for analyzing change trajectories (i.e., change from session to session) that we plan to present in a paper to come.

5.3 Recruitment and procedures

Within the frame of regular client recruitment to the sites, clients were asked to participate in the study. After obtaining informed consent, they were included in the study. Those clients that did not participate in the study were given treatment just as usual. In the Pilot Study, the clients were asked to complete the STIC initial either before the first session, or just after. However, some completed the questionnaires later in the treatment in the Pilot study.

In the Pilot study, STIC Initial was completed at the start, and STIC intersession was completed subsequently before every next session, and no outcome package was included. In the RCT, the clients completed the STIC Initial and an additional outcome package before the first session, and the STIC Initial and outcome package at termination. Clients that did not fill out the final questionnaires within a given time limit were sent SMS or email notifications.

The clients randomized to using STIC in therapy completed the STIC Intersession before every session after the first and until the last. In Paper 3, we used pre and post data for the clients who were randomized to use STIC Intersession. As mentioned, the data from session to session were not included in this study.

All questionnaires were completed by clients electronically via secure zone Internet and could be filled out from any computer, tablet or cellular phone with Internet access. The Psychotherapychange.org website received, analyzed and securely stored client data and feed them back to therapists on the therapist website.

5.4 Measurements

The RCT study collected data using several outcome measures, while the Pilot study only used the STIC. In Paper 1 and Paper 2 only the STIC measure was applied. All clients included in Paper 3 had used STIC as part of their treatment but none of the STIC measures were used in the analysis. Four other measures were used and I will now briefly present each of these measures.

5.4.1 Individual distress – Level of depression

One common individual distress is depressive symptoms, and the Beck Depression Inventory (BDI; (Beck, Steer, & Brown, 1995) was chosen as the questionnaire to measure its levels and change during treatment. BDI is a widely used 21-item questionnaire assessing symptoms of depression and variation over time. The sum-score expresses the depth of the depression, graded from no clinical depression (0 – 9), through mild (10 – 19), moderate (20 – 29), and severe depression (30 – 63). BDI has a clinical cut-off level of depression at 14.29. In the sample used in Paper 3 the Cronbach's alpha at pre-treatment was .93 (CI: .91 - .95), and it was .95 (CI: .94 – 96) at post-treatment.

5.4.2 Relational – first level – Couple distress

The first level of relationship that was assessed in this study was the couple relationship. To measure satisfaction and distress at this relational level, the Revised Dyadic Adjustment Scale (Busby, Christensen, Crane, & Larson, 1995) was chosen. This questionnaire is a widely used 14-item valid and reliable tool providing a global measure of each partner's assessed couple consensus, satisfaction, and cohesion toward the spouse. The scoring range is 0 - 69, with

higher scores representing better adjustment. The clinical cut-off is 47. Cronbach's alpha at pre-treatment for the sample used in Paper 3 was .85 (CI: .82 - .89), and at post-treatment, it was .88 (CI .82 - .91).

5.4.3 Relational – second level – Family distress

The second level of relationship included in this study is the family. Family functioning was measured by the general functioning subscale of the Family Assessment Device (FAD; (Epstein, Baldwin, & Bishop, 1983). The FAD is a widely used 12-item questionnaire as a brief method of assessing overall family functioning. The scale range is between 1– 4, with a total score from 12 - 48 (lower is better). FAD has a clinical cut-off at 2.00. In the sample used in Paper 3, the Cronbach's alpha at pre-treatment was .89 (CI .86 - .91), and at post-treatment, it was .91 (CI: .88 – .94).

5.4.4 Societal functioning – Work functioning

To measure societal functioning, we chose to use one of the subscales from the quality of life measure SF-36v2 Health Survey Standard Version (2004) (Trust, 2016). SF-36v2 is an 11 item instrument that measures the quality of life that is internationally and cross-culturally comparable within areas of physical and mental health, activities, and well-being. Only one of the subscales was used in this study - the role emotional scale - that measures work functioning concerning emotional challenges. In the sample used in Paper 1 the Cronbach's alphas for the role emotional scale were .91 (95% confidence interval CI: .88 - .93) at pre-treatment and .92 (CI: .88 - .94) at post-treatment. No clinical cut-off was established. Hence, we defined in Paper 3 a cut-off level for this subscale. As a prerequisite in identifying the clinical cut-off for the SF-RE, we used a normative sample of the Norwegian population (Loge & Kaasa, 1998) even if this was done using the first version of SF. We do so since international normative samples have compared the version 1 and 2 and found the norms to be interchangeable (Jenkinson, Stewart-Brown, Petersen, & Paice, 1999; Morfeld, Bullinger, Nantke, & Brähler, 2005). Since the mean age of our sample was 40,6 (±9,2) we used the mean for the normative sample for those from 30 to 49 years old, 85,4 (±29,4).

5.4.5 Systemic Therapy Inventory of Change – STIC

As previously mentioned, the ROM chosen for this study is the STIC. The factor structure of the STIC System scales is empirically derived, atheoretical, and generic by design so they can be used across different theoretical models, clinical populations, and contexts, making it suitable for comparative effectiveness research. As mentioned, this ROM system is multi-systemic (assessing information at the individual, couple, and family levels) and multi-dimensional (collects information about specific aspects of a client's life at each system level). The STIC is found to have good internal validity (Pinsof et al., 2009) with a Cronbach's alpha as high as 0.94 for the different subscales. The system scales in STIC have been found to be reliable (Pinsof et al., 2015b) and have strong concurrent validity with well-validated and widely used criterion measures (Pinsof et al., 2009). The factor structure of the STIC scales has been replicated and normed on a second clinical sample and a "normal" random representative sample of the US population (Pinsof et al., 2015b).

In Paper 1, we made use of one of the scales, the individual problems, and strengths scale (IPS). This scale had a Cronbach's alpha at .89 in the total sample. In this paper, we tested how much a complex measure can be simplified before it loses its ability to explain the clients' lives in a meaningful way. Some might question the choice of an individual measure in a study within CFT when relational measures were available. There are two reasons for this choice. The first is that all clients are answering to this scale, no matter if they are young or old, in a current relationship or not. The consequence is that we obtained a high response rate with data from all 841 clients. None of the other scales obtained answers from all the clients. The second reason is that sometimes, within CFT, only the relational aspects are given focus and attention and the individual aspects can, therefore, be ignored. Since relational aspects are given focus in Paper 2 and 3 we, therefore, focused on responses to the individual problems and strength in Paper 1.

5.4.6 Domestic violence

In Paper 2 we used several scales from the STIC to investigate physical couple and family violence. Three items in the STIC directly measure physical violence. The Relationship with partner scale in the STIC has one item addressing physical violence between the members of the couple: "get into shoving or hitting each

other when we fight." In the paper, we name this as Couple violence. The family-household-scale has two items addressing physical violence exerted within the family: "Someone in my family is physically abusive to other family members," and "There is someone in my family who pushes other family members around physically to get his or her way." We combined these two items and named it Family violence in the analysis presented in Paper 2.

5.5 Summary of the method, data, and sample

Table 1 presents a summary of the three papers regarding the domain, focus, the main type of analysis, design, data type, and sample.

Table 1.
Overview of the domain, type of data and sample for the three papers

	Paper 1	**Paper 2**	**Paper 3**
Domain	Individual (and general)	Relational – Couple and family	Societal
Focus	Simplification of measures	Physical Violence	Work functioning
Main type of Analysis	CFA	SEM	Hierarchical regression
Design	Cross-sectional	Cross-sectional	Longitudinal
Data type	Pre-treatment	Pre-treatment	Pre- and post-treatment
Sample	All Pilot and RCT data	Pilot and RCT data, age > 18	One branch of RCT data
Sample size	841	830	165

6. Summary of results

I will now present a summary of the results for each of the papers before I, in the next chapter discuss the main findings and the methodological considerations.

6.1 Paper 1

Monitoring therapeutic change through diversity or simplicity? A conservative, critical Confirmatory Factor Analysis test of a Routine Outcome Monitoring system.

This paper aimed to assess how much simplification and reduction can be conducted on a comprehensive Routine Outcome Monitoring system before the system no longer helps the therapist to understand the clients and the therapeutic process.

We tested this on one of the scales of the STIC system, and the results showed that the diverse model, with sub-scales, had good model fit, good variation of factor loadings, variable covariances, and a theoretical rationale. None of this was found on the alternative models. The results show that when using a simplified, omnibus model, there is an increased risk of misunderstanding clients and their therapeutic experience. Possible implications are that clinicians and researches should avoid using a simplified measurement scale in their clinical evaluation of the severity of problems or change in therapy if they have access to a more comprehensive, refined one.

6.2 Paper 2

Physical couple and family violence among clients seeking therapy: identifiers and predictors.

The aim of Paper 2 was to identify and describe the group of clients that had issues of physical couple and family violence, and to develop a model that can help to discover physical violence and help therapists to assess what actions to take in therapy to prevent further physical violence.

The results were that one-in-five clients experienced physical couple violence in their current relationship, and one-in-four experienced physical family violence. The group of clients that experienced couple violence was characterized by having a lower income, more prior experience with psychotherapy, more experience with alcohol abuse in childhood, and far more physical family

violence in their current family compared to those without such experiences. Our model predicting physical couple and family violence had a very good fit. The model explained 53% of Physical violence and had three positive predictors (expectation, anger, and sexual satisfaction) and one negative predictor (self-control). Contrary to our expectation, sexual satisfaction was found to be a positive and not a negative predictor. Another unexpected result was that alcohol abuse was not found to be a predictor that significantly improved the model.

6.3 Paper 3

Change in work functioning from pre- to post-treatment in feedback-informed couple and family therapy in Norway.

The primary aim of Paper 3 was to analyze the change in work functioning from pre- to post-treatment, concerning distress on individual, couple and family measures. The group of clients improved from start to end of treatment on work functioning as well as on the individual-, couple- and family measures with medium to high effect sizes. When each case was analyzed individually for change on work functioning with the clinical significance criteria, the results was that 54 % of the clients had improved or even recovered.

We found that the level of depressive symptoms, couple distress, and family functioning predicted work functioning at pre-treatment. The level of work functioning was significantly better predicted when the relational measure of couple distress was included, compared to when only the individual measure was used. The prediction further improved significantly when the next level of relation, the family measure, was included.

None of these measures' levels (i.e., individual, couple, or family measures) at pre-treatment predicted work functioning at post-treatment. However, change on these measures from pre- to post-treatment predicted 53 % of work functioning at post-treatment, when controlling for the level of work functioning at pre-treatment.

7 Discussion

This section starts by discussing the main findings for each of the three papers concerning established research and relevant theories. Second, I present a general discussion of study design and generalization of results. Finally, I discuss strengths and limitations related to each of the three papers.

7.1 Discussion of main findings

7.1.1 Paper 1 – Simplification of measures

Regarding Paper 1, I discuss the results concerning the ongoing debate on what measures to use when assessing levels of distress and change in therapy and, more specifically in the discussion of the use of ROM in therapy.

Explicitly, Paper 1 investigates the question how to measure factors related to what is reliable, meaningful and helpful. If change is to be measured we have to be sure that the measure is not only sensitive to change and has a high level of specificity, but that the measured change has clinical significance (e.g., through the Jacobson and Truax (1991) formula) and is meaningful (Comer & Kendall, 2013; Langkaas et al., 2018). We should also make sure that the measured change is a change in the aspects of people's lives. It is aimed at targeting. We found that the IPS sub-scales model has support for measuring what it intends and that it in its construct is meaningful both for researchers and clinicians, while the IPS total scale did not satisfy these criteria.

The background for addressing this topic in my research is because several measures come in different versions. One example is the SCORE measure that is available in versions of 40, 28 and 15 questions (Carr & Stratton, 2017; O'Hanrahan et al., 2017). The more questions, the more details, but it is time-consuming for clients to answer a large number of questions and it might be overwhelming for therapists to investigate all the questions and answers. Hence, the shorter versions are often selected. Anyhow, the replies are most often combined to subscales and total scales before they are fed back to the therapist for clinical purpose, or analyses of the data are being performed for research purpose. The contribution to the scientific debate in Paper 1 is that even though it is less time-consuming to use the shorter measures or to simplify conceptually different items into one total scale, we argue that this should not be done.

However, Miller et al. (2006) suggest that asking only a few questions to the clients is enough to track meaningful change and to assess outcomes. Their claim has been empirically supported for many years (see e.g., Anker et al., 2009; Owen, Duncan, Reese, Anker, & Sparks, 2014). However, others argue that more detailed questions will provide therapists with more useful information (Zinbarg et al., 2018), and for this reason they suggest using a more complex ROM (i.e., the STIC system). Indeed, the Paper 1 study has empirically tested if a complex ROM can be simplified to the level of those one item ROM's that are available. We investigated this in Paper 1 on one of the STIC scales. Even if the conclusion was clear for this scale the same type of analysis should be performed on all the STIC scales and on other ROM systems to see if the results are consistent. The reason not to analyze all STIC scales in this study was that it was not found to be feasible within the scope of this study.

However, it is important to notice that this discussion, in particular, applies to comprehensive measures such as the STIC that aims to cover several distinctively different concepts in one measure and then simplifying these into an omnibus model. Specified measures targeting one area of people's lives, e.g., anxiety, are possibly not affected by the same problems when calculating total scores as a complex measure as the STIC.

In summary, Paper 1 contributes to the main aim of this thesis by assessing one dimension of how to investigate couple relationships and change, namely what level of simplification to use on a comprehensive measure.

7.1.2 Paper 2 – Domestic violence

Paper 2 investigates domestic violence as an example of a relational measure. When we are to investigate a complex interaction in a dyadic relationship, like domestic violence, there are several aspects to consider. I want to address three of those. First, there is a need for a definition of the phenomenon I want to investigate. Secondly, this phenomenon has to be placed in a theoretical framework. Third, since the literature on domestic violence has established a strong connection between alcohol abuse and violence, and our results did not support that theory, I close with a discussion of alcohol abuse and its relation to domestic violence.

7.1.2.1 Definitional considerations

In most studies on the general population, the question is if they have experienced psychological, physical, or sexual violence in a relationship. This approach is different than asking specifically about physical violence, as we did in this study. This study is different than most studies on general populations also when it comes to timeframe of the experienced violence. In most studies on the general population the question is if they have experienced violence in any relationship until now. In this study, we focused on physical violence in the current relationship, not any possible prior relationships. Considering these differences, when comparing the results from this study, we concluded that physical violence has much higher prevalence in a clinical CFT sample than in the general population. This finding is in agreement with Jose and O'Leary (2009) and Stith et al. (2011).

As mentioned, in the Paper 2 study, 20 % of the clients describe that physical violence is part of their current relationship. I learned from the two national studies on the general population in Norway that 5% had experienced physical abuse in the last 12 months (Haaland et al., 2005) and 14-16% in their lifetime (Thoresen & Hjemdal, 2014).

However, since a meager percentage of the cases in CFT are being reported for violence (see chapter 3.6 for a discussion) it seems as if most of the violence among clients in clinical samples continues as uncovered even in CFT treatment. Based on the high prevalence of physical couple and family violence found in this study seen in contrast to the low portion of reported violence in BUP systems (Reigstad et al., 2006), the use of standardized screening of clients has been suggested (see, e.g. Middelborg & Samoilow, 2014; Stith et al., 2011), and this is something I also recommend.

7.1.2.2 Theoretical considerations

Among a variety of relevant psychological theories on physical violence, we based our model on the conceptual theory of Gausel and Leach (2011) that explains how people may react to moral failure. Moral failures, like forgetting important appointments or shameful emotions are not easy to hide from the others in a family. Hence, the chance of being condemned for moral failures is present. Even more so with illegal actions like using violence. One central

component of the Gausel and Leach (2011) theory is how a person appraises his or her moral failures. What we found was that the more a person experienced that the other partner expected more of him/her than he could live up to, the more physical violence was reported. This result is in line with the contextual model of Gausel and Leach (2011) explaining that violence can be an action to protect their social image (others' evaluations of them) as there is a chance that the other person might evaluate them as persons who are not responsible or maybe not even able to fulfill what is expected. This feeling is associated with threatening their social image and possibly also their view of self. If this is appraised as a specific flaw or defect, they will, according to the conceptual theory, try to reconcile and become better. However, if it is appraised as a global defect (i.e., "I am bad") it will lead to withdrawal or attack. None of those reactions are positive for a close relationship (Gottman & Silver, 2015), but attacking the other with physical violence was the one we found associated with expectation in Paper 2. Further, anger towards the partner was also found to be a significant predictor of physical abuse. In other words, the more anger, the more reported violence. Anger and expectation are connected to the situations the couples are in when the violence is exerted. Contradictory to this suggestion, the simple model of domestic violence by Cano et al. (1998) and the more complex model of DeMaris and colleagues (2003) include aspects from the client's background. We tested if such background factors would positively contribute to our model. But, they did not. This fact suggests that background factors might not be as important as suggested by those theories. The reciprocal, situational violence are predominantly explained by the situation the couple is in at the time of the violent acts. This theory is strengthened further by our investigation of experiences of alcohol abuse in their family of origin.

7.1.2.3 Alcohol abuse

As many as 41% (339) of the clients reported that someone in their family of origin had used too much alcohol. Since it is found that such experiences increase the risk of developing problems in need of psychotherapy treatment (Dube et al., 2001) one may speculate whether negative experiences in the use of alcohol in their family of origin had some association with their own need to seek therapy later in life. This factor is an empirical question that should be scrutinized, and with the available STIC data, it is possible to do so in the future.

Almost 20 % of the clients reported that they had been using too much alcohol themselves; further that 10 % of their partners abused alcohol. However, this study found that alcohol abuse was a less important predictor of physical violence than the literature suggests. For instance, prior studies have determined that physical couple violence is associated with alcohol and substance abuse, both at present and in adults' family of origin (e.g., Coker et al., 2000; Fals-Stewart, 2003; Fals-Stewart et al., 2005; Stith et al., 2012).
Even if such significant bivariate correlations also were found in our sample, they did not contribute to our model when it comes to the explained variance. Furthermore, the alcohol abuse variables decreased the model fit statistics. However, this does not mean that these variables do not predict physical couple and family violence, just that they do not contribute enough to improve the model used in our analysis.

In summary: Paper 2 contributes to the main aim of this study by investigating one specific part of couple relationships among clients in CFT treatment, namely physical violence and discussing the findings up against common theories and the literature on domestic violence

7.1.3 Paper 3 – Work functioning

Paper 3 investigates work functioning as an example of a societal measure. I start by addressing work functioning as an outcome measure in CFT. Second, I discuss how this focus made it possible to test and possibly challenge systemic theory. Third, I discuss the results from the Paper 3 study in light of what subsamples to use in CFT - and psychotherapy research. Finally, I discuss how nomothetic and idiographic knowledge can be applied in clinical practice.

7.1.3.1 Work functioning as an outcome measure

Paper 3 focuses on clients' work functioning. However, since most clients are motivated for therapy based on their individual and relational distress, it is understandable that those aspects are of the highest importance to them. It is, therefore, not strange that most therapists and researchers have had the same focus. Indeed, what is important to the client should be of importance for therapists and researchers (Anderson, 2019; Ulvestad, 2007). These individual and relational aspects are also what have been the focus of most studies of

change in CFT (Sexton et al., 2013). However, researchers within CFT, such as Crane (Crane, 2011) has argued that we also have to investigate aspects of great interest for stakeholders and society in general, such as cost-effectiveness of the treatment. Another example of great importance for society is work functioning. Even if it has significance for the individual and the society it has rarely been used as an outcome measure in psychotherapy (Ogles, 2013), and to my knowledge, this paper is original in its focus on work functioning within the area of CFT research.

7.1.3.2 Systemic theory

Johnsen and Torsteinsson (2012) state that it has been difficult to assess the theoretical building blocks that lie behind the therapeutic interventions in CFT. One of these building blocks is the systemic theory (Hårtveit & Jensen, 2004), stating that all persons and actions are interrelated and that to understand what is going on we have to expand the context of the phenomenon instead of going closer and scrutinize it. In other words, this implies looking at the background and its relationships. An interesting by-product of the analysis in this Paper 3 is that we tested if the systemic theory had empirical support in one area of people's lives, i.e., work functioning. We tested this by analyzing predictive models with different levels of context. Work functioning is an individual measurement of a societal measure, and the question we investigated was if we got better understanding of the clients' work functioning when we added levels of context. Since the prediction of work functioning was significantly improved for each step of the analysis, from the individual to couple, and then next to family, this implies that to understand a phenomenon, such as work functioning, therapists and researchers alike will gain best understanding by widening the scope of focus and include also the relational aspects of their clients' lives. Indeed, this is a result that supports the essence of systemic theory. Clinically it implies that therapists should attend to the client's difficulties not only at an individual level but also at the levels of intimate relationships (i.e., couple and family levels).

However, our results also challenge the common assumption that it does not matter where to make a change in a systemic system (Hårtveit & Jensen, 2004; Johnsen & Torsteinsson, 2012), because the correlation analysis identified that some of the variances in work functioning were explained by our individual

measure, and other parts of the variance were explained by the couple and family measures. This fact implies that it might matter what to focus on and when. I find this in agreement with the two CFT therapists Pinsof and Lebow (2005), who suggested what they called "differential causality" explaining that different variables in the systemic interaction may vary in strength, distance, and impact. A consequence of this differential causality is that it matters what to focus on in therapy at which time-point. In Paper 3 we identified the individual measure (depressive symptoms) as the most influential on work functioning when compared to the relational measures (couple distress and family functioning). This result suggests that targeting depressive symptoms would be an appropriate focus at the start of treatment if improved work functioning was the primary aim of the treatment. The suggestion to focus on depression early in CFT in treatment is in line with Tilden, Theisen, Wampold, Johnson, and Hoffart (2019a). They suggest targeting depressive symptoms because it predicted relationship improvement both at the end of treatment and three-year follow up. Since our study found that both depressive symptoms and relationship improvement predicted better work functioning, suggesting focusing on depression early in treatment seems reasonable. However, clear conclusions cannot be drawn from a single study like this, but the mentioned result calls for further studies where this systemic theory is empirically tested.

7.1.3.3 Subsamples when assessing change

This study identified improvements from pre- to post-treatment on all the measures included in the study. When analyzing the total sample, the effect size was small to medium due to the many clients initially in the non-clinical range of the measures and thus unable to improve much on those measures. This issue is important when it comes to the choice of measures, but also on the inclusion of clients in the analysis. If measures were used that better respond to change in a non-clinical sample the effect on such a sample possibly could be detected. However, our results when using the total sample were in line with the review of Sexton et al. (2013), showing low to medium effect sizes of CFT treatments. In contrast to when analyzing the total sample we identified a medium too high effect size (d = .71 - .84) when analyzing the clinical subsamples in Paper 3. This is higher, or comparable to effect sizes for specialized treatments (e.g., for depression (Cuijpers, van Straten, Bohlmeijer, Hollon, & Andersson, 2010; Klein, Jacobs, & Reinecke, 2007), anxiety (Acarturk, Cuijpers, van Straten, & de

Graaf, 2009; Cuijpers et al., 2014), eating disorder (Vocks et al., 2010; Wilson, Wilfley, Agras, & Bryson, 2010)). Since the treatment provided in this study was general CFT and not specialized at any of those measures this is a promising finding for CFT therapy. When compared to the results of the Sexton et al. (2013) review of meta-analysis within CFT (d = .22 - .54) our results are considerably higher. However, this discrepancy leads to an important aspect when analyzing change in CFT samples: The question of what sample to include in power analyses. This factor is particularly important when comparing to other types of treatment, such as individual and group treatments. The analyses in this study illustrated the importance of analyzing clinical subsamples in CFT when comparing effect sizes.

7.1.3.4 The contrast between nomothetic and idiographic information

It is important to notice that even when the general effect size was high for the clinical sample in Paper 3, not all clients improved. Almost 50 % had no change or deterioration, a finding that is common in psychotherapy research (Ogles, 2013). However, consciousness about the discrepancy between the results from the group change (i.e., nomothetic knowledge) and the individual change (i.e. idiographic knowledge) is therefore crucial for therapists. First, the discrepancy between the results highlights the fact that there is a distribution of outcomes when calculating on an individual level within the sample of clients, something that is masked when presenting a point estimate of effect size on a group level. Even if a statistically skilled reader will be able to see the distribution through the means and standard deviations of the variables used to calculate the effect sizes, this is often not the case for practitioners. Indeed, presenting effect size and clinical significance *together* clarifies this and also makes it easier to understand and to apply to clinical practice. The common use of presenting effect sizes without presenting clinical significance might explain the mentioned skepticism among some CFT therapists towards quantitative studies (Pinsof & Wynne, 2000; Sundet, 2019; Tilden et al., 2010). In other words, since the therapist meets every single client one by one, quantitative research is often perceived as something strange and not clinically useful by the therapists as long as this is interpreted as results achieved on a group level. Second, knowledge of one treatment that is found to be helpful for a diverse group of clients should help therapists to make therapeutic decisions on what treatment or intervention to suggest or provide. However, we do not know yet if the suggested approach is

beneficial for this single client. Thus, knowledge from the analysis at the individual level can help therapists to be humble in the use of those methods as these might not be helpful for the present clients in therapy. The statistically more aware therapist knows that nomothetic knowledge is based on groups of clients and that the conclusions are based on the majority of the sample and that the conclusion would not be the same for every client in the sample distribution. My opinion is that both nomothetic and idiographic information is useful for clinical practice, and maybe knowledge of the contrast between the two types of information is of most importance when trying to tailor the best possible treatment for each client.

In summary: Paper 3 contributes to the main aim of this study by investigating change from pre- to post, treatment on the societal measure, work functioning, and assess that concerning change on couple relationships and family functioning. This focus made it possible to assess systemic theory that forms the ground for CFT treatment. Finally, in this discussion I discussed how to apply knowledge gained from a thesis like this on couple relationships into a clinical setting.

7.2 General discussion of the method

In this chapter, I will start by presenting a general discussion of study design before I discuss reliability and validity.

7.2.1 General discussion of study design

The choices made in this study on what, who, when, and how to measure has inevitably impacted the results of the study. I chose to focus on two aspects when investigating couple relationships within CFT, i.e., physical violence and work functioning. Ideally, a research project should start with a research question followed by a decision of the best method to answer that question (Comer & Kendall, 2013). However, that was not the case in this study, as it builds on the data from a multi-site pilot (Tilden et al., 2015) and an RCT-study (Tilden et al., 2019b). The choice of measurements was made, and data collection was already in progress as this study evolved. Hence, the research questions arose from a combination of the data that was being collected and my interest as a researcher. Since the measures already were set this limited the questions possible to ask. I could not ask questions that were impossible to answer through the measures

previously being collected. However, it is possible to argue that having several measures predefined also inspired and broadened the creative process of deciding research questions. Indeed, several measures were included, and a wide variety of interesting questions were possible to investigate. The rich data open for investigation of anxiety, depression, cost-benefit, therapist differences in outcome, process analysis of treatments, network analysis, and many other aspects. Guided by interests as a therapist and researcher the specific research questions were chosen, but not without the impact of funding possibilities and literature research to see if the questions already had been thoroughly answered. If the order had been ideal, I would have decided the research questions first and then chosen between qualitative or quantitative methods or both to answer those questions. Since the choice of measures was set, I was not tempted to develop my own measures and since commonly used measures were chosen it made the results of this study more comparable to other studies.

7.2.2 Discussion of reliability and validity

Validity relates to whether we are measuring what we intended to measure and if the evidence supports the inferences to be accurate or correct (Drost, 2011). Reliability, on the other hand, is a measure of the stability of a measurement over a variety of conditions or time in which the same results should be obtained (Drost, 2011). I will start by discussing validity, and according to Drost (2011), there are four types of validity that researchers should consider: statistical conclusion validity, internal validity, construct validity and external validity. I will in the following exemplify possible threats to these types of validity in this study. (1) Statistical conclusion validity can be challenged by low statistical power (i.e., an actual difference is not recognized because of low power). In paper 3, this issue limited the analyses to be performed. Because of low N if the sample was divided on the different sites this could not be performed. Heterogeneity of respondents is also an issue that could threaten the validity. For example, when assessing change in a CFT sample heterogeneity of distress among the two in the couple is an issue. Thus, in Paper 3 we presented the results both on the total sample and on the clinical-sample. (2) Internal validity could be threatened by (a) the missing data (discussed in chapter 7.3.4) and (b) that the change measured in Paper 3 stems from other historical events happening outside therapy context. (3) Construct validity could be threatened because we only used

self-assessment measures. Even if the clients are closest to their own life (Anderson, 2019), something that strengthens the validity of the study, client self-reports could be biased (Comer & Kendall, 2013) and it would have increased the construct validity if we used a multi-informant strategy (Solem, 2002). (4) External validity regards the generalizability of the results, which could be threatened because of the uniqueness of the Norwegian sample.

This study included a reasonably high number of participants (N=841) that vary in many aspects. The clients vary in age, levels of distress, prior experience of therapy, and reason for attending treatment. Because the participants were recruited from regular intake to these treatment sites, one should assume that the participants represent the variety of the client population seeking CFT, and hence, this heterogeneity strengthens the external validity of the results. Because this study was conducted within the three CFT levels offered by the Norwegian welfare system, we suggest that the sample is representative of the Norwegian CFT population, hence supporting the external validity of the findings. Due to the Norwegian tax system that offers public insurance for every citizen seeking public professional help, therapy at the first level of care (e.g., family centers and family offices) is free of charge and only a low deductible has to be paid at the other two levels (i.e. specialized treatment offered in public and private hospitals). This information implies that clients from all levels of society have equal access to the treatments offered, not only the economically benefitted. For this reason, I assume that the participants in this study reflect the general population in Norway considered from a social-economic/social class perspective. This phenomenon strengthens the generalizability to clinical contexts similar to the Norwegian system, however it weakens the generalizability compared to clinical settings that differ.

There are two additional aspects connected to the fact that this study was conducted using only a Norwegian sample that I also want to address. Firstly, in Norway, there are no national restrictions on treatment length and very few instructions on the choice of treatment models. Hence the therapists have a great deal of freedom that might lead to more diverse treatments than would be possible in countries with stronger restrictions. Indeed, this might lead to less globally generalizable findings. Secondly, some may argue that the sample is not representative of an international population since all participants come from

Norway. However, Norwegians seem to be similar to other nationalities in the Western world (Becker et al., 2018). For example, the divorce rate in Norway is similar to citizens from other European countries (Eurostat, 2015). Further, it is likely that individual and relational problems are shared internationally, and if differences occur, they are not as expected (Owe et al., 2013; Vignoles et al., 2016). Moreover, the four agencies participating in this study were located in different parts of the country and thereby included clients from different socio-economic groups of the population.

Concerning the therapists and the therapy being offered there are some considerations I will address. It strengthens the external validity of the results that these data stem from ordinary practice with therapists with different levels of training, theoretical influences, and years of experience as this should reflect the variety of therapists who conduct CFT in Norway. Because most studies that are including data from RCT design are carried out in university settings with (student) therapists under strict supervision (Comer & Kendall, 2013), the aspects as mentioned earlier in this study demonstrate that this study differs from the majority of studies by being conducted closer to natural treatment contexts.

I will now discuss two aspects of reliability that relates to this thesis.
The first relates to if the measures we used were reliable. The second relates to the reliability when only using the clients themselves to assess these measures. The reliability of the IPS scale of the STIC system was thoroughly examined in Paper 1, and the results indicated that the constructs of the subscales model were good. These results are in line with the results from an analysis of the STIC on a representative US sample (Pinsof et al., 2015b) a factor I consider supports the reliability of this thesis. The Paper 2 study also used STIC measures, and the constructs were found to be good. In Paper 3, other measures than STIC were used, but the Cronbach's alpha's on the included measures in our sample ranged from .84 to .95. This result demonstrates a satisfactory level. Thus, the reliability of the measures was assessed in all papers and found to be good.

However, a single-informant strategy for data collection was chosen for this thesis something that could challenge the reliability and, as mentioned, the validity. Therapist evaluations and evaluations from others within the family or concerning them could then have been used to assess the levels and change on individual and relational levels. In the Paper 3 study, we could have asked

employers to rate the clients' work functioning pre- and post-treatment. In the analysis we could thus have compared the scores from self-evaluations with the outside rater's scores. If they indicated the same change as the clients themselves this would have increased the reliability of the study. If they were different it could have led to interesting discussions. However, with regards to Paper 3 it would also have made an already complicated paper even more complex. Another multi-informant strategy could be to use outside raters who met the clients in separate assessment sessions and then comparing the results with the self-evaluations. This option would have added extra sessions that would have been costly and taken extra time from the clients in the study. Indeed, it would have challenged the ideal of ordinary practice that was incorporated into the RCT study. This approach could also have challenged the idea of valuing the clients' view higher than the experts because outside raters easily could be defined as objective raters in contrast to the client's self-evaluations. Indeed, Pinsof (2017), the developer of the STIC system, declares that it is the client's narratives that should have the highest credibility. However, the clients' narratives about a particular score are hard to get access to outside a therapeutic setting where the STIC profiles are discussed with the client.

7.3 Discussion of strengths and limitations

In this discussion of strengths and limitations I will address aspects related to each of the three papers individually before I address missing data. The focus on missing data is because it is an issue often not thoroughly examined (Wood, White, & Thompson, 2004), it applies to all papers, even if it is especially important for the analyses of the time series data in Paper 3.

7.3.1 Paper 1

Paper 1 investigated the simplification of measures, and in this discussion of strengths and limitations, I chose to focus on two aspects. That is generalization of results due to the choice of scale and the choice of method used to analyze and assess the data.

7.3.1.1 Strengths and limitations of generalization due to the choice of scale to be analyzed

In Paper 1, we chose to investigate the individual scale of STIC, even if relational scales could be considered as closer to the core of this thesis addressing

CFT. We had the option to do so but chose not to due to the following reasons. IPS is the scale most easily generalizable to other measures due to its wide spectrum of areas of interest related to the individual client (e.g., substance abuse, disinhibition, and life functioning). Not so many measures cover experiences of the family of origin, couple relationship, family functioning and relation to children as covering individual functioning (Ogles, 2013). Moreover, those relational measures are more restricted in their areas of interest than the IPS scale. It is also true that the individual scale of STIC had the strongest power (i.e., highest N), and since the plan is to analyze the other scales in the STIC system in coming papers it seemed obvious to start by analyzing the individual scale.

What we gained by choosing the STIC Initial, and thus strengths to this study, is that the results can be generalizable to other measures used as an initial screening of clients (e.g., OQ-45; Lambert et al., 1996) that have found to significantly correlate to STIC scales (Pinsof et al., 2009; Zinbarg et al., 2018) and that we could use the total sample of 841 clients. However, it is important to notice that Paper 1 investigated STIC in a more general discussion of ROMs, where STIC is only one of many different ROM systems. Some of the other ROM systems are less comprehensive and do not cover such a diverse set of items as STIC and might therefore not suffer from the problems concerning simplification of the scales, as identified in Paper 1. Hence, a possible limitation of Paper 3 is that only one of the subscales of one out of several ROM's were analyzed.
I, therefore, suggest analyzing other ROM systems in similar ways as performed in this study.

7.3.1.2 Strengths due to the choice of method of analysis

Paper 1 used Structural Equation Modeling within Amos (Blunch, 2013) to run the Confirmatory Factor Analysis. I consider this to be a strength to the study on three aspects when compared to run the CFA from SPSS. First, Confirmatory Factor Analysis in SPSS does not take correlation of the factors into account (Kline, 2016; Tabachnick & Fidell, 2014). Secondly, in SPSS, the factors are not subjects in themselves, as they are in Structural Equation Modeling (Kline, 2016). In Structural Equation Modeling the factors are considered as latent variables, and since they all are allowed to correlate, the factor loadings to each item will be affected by this correlation. With fewer factors, the correlation between them will not have such an impact. Finally, the advantage of using

Structural Equation Modeling when running Confirmatory Factor Analysis is that we can pre-define all the factors and the items each of them are theorized to load into (Kline, 2016).

For researchers like me who are visual learners (Felder & Spurlin, 2005), it is also an advantage in using Structural Equation Modeling within Amos to run the Confirmatory Factor Analysis because of the model and its estimates are displayed in figures and tables, not only in tables as in SPSS.

7.3.2 Paper 2

Paper 2 investigated physical violence among clients in CFT, and I choose to focus on three aspects of strengths and limitations concerning this study. The first regard the definition of the violence that was investigated. The second the items used to assess, and the final paragraph discusses the choice of emotion used in the model.

7.3.2.1 Strengths and limitations due to the strict definition of physical violence

It is a strength to this study that the type of violence is specified as physical violence, and that the reports are limited to the current relationship. However, the strict definition could also be considered a limitation because emotional violence has been found to be highly prevalent in clinical samples (Jose & O'Leary, 2009). Another possible limitation is that the clients did not report who exercised the couple violence: whether it is themselves as perpetrators, whether they are exposed to it by their partner, or whether it is both. However, most of the violence detected in CFT samples are most probably reciprocal, i.e., situational (Fusco & Fantuzzo, 2009; Simpson et al., 2007; Stith et al., 2011). Further, recent research on violent conflicts suggests that both parties prefer to construe themselves as victims instead of perpetrators, even if both exercised violence (see, e.g., Gausel et al., 2018; Mazziotta, Feuchte, Gausel, & Nadler, 2014). Hence, it is not unthinkable that if the clients were asked to indicate who exercised the violence, there is a high probability that they indicated the other as the perpetrator and themselves as the victims. In contrast, Karakurt et al. (2016) identified several studies that reported male-to-female violence and female-to-male violence. Future studies could incorporate these two different measures and assess if females indicate more male-to-female violence and males more female-to-male violence.

Another limitation is that there is no differentiation between severe physical violence and minor physical violence in this study. However, one could argue that by dividing the sample into three groups rather than two, we could gain more information: 1 = no violence, 2 = seldom violence, 3 = often or more. However, we had two reasons for not doing so. First, even if violence seldom occurs, we do not know how severe it is, and a few very severe incidents might be as powerful as many less severe incidents. Second, if violence only rarely occurs it can still be an important part of their family life (Middelborg & Samoilow, 2014).

7.3.2.2 Possible limitations due to the item used to assess couple violence

Someone might question the validity of the item used to assess couple violence because this item is ambiguous whether the clients indicate *shoving* or *hitting*. It could thereby be presented as a limitation. However, from a clinical point of view, it makes sense to have both categories included because the point is that physical force has been used. From a clinical perspective, it is not the distinction between shoving and hitting that is most important, but the fact that someone inappropriately has used force towards the person. However, from a research point of view, it would be better if we knew what part of the question they were answering. Have they been shoved or hit or possibly both? In the case of STIC, the founders of the system have chosen what was considered clinically useful information that could be explored in more depth in the following therapy session. Clinically, the questionnaire and the replies are thus in this system considered as starting points for further conversation (Pinsof, 2017). Other possible limitations are that couple violence was only captured with one single item and further that we only assessed cross-sectional data and therefore could not assess whether the couple violence was reduced over therapy.

7.3.2.3 Limitation due to number of emotions included in the analysis

In Paper 3 we used a narrow focus on anger as the only emotion identified as a predictor of family violence in our model. Undoubtedly, there are other emotions, such as a feeling of rejection (Gausel, 2013; Leary, Twenge, & Quinlivan, 2006) or of hate (Staub, 2005) that are found to predict violence and anti-social behavior. Among these feelings, only anger is explicitly monitored through the STIC relationship with the partner scale and was, therefore, the feeling included in the analyses. This situation is a limitation.

7.3.3 Paper 3

Paper 3 investigated work functioning, and in this discussion of strengths and limitations, I will discuss issues concerning the choice of sample, adherence to protocol and issues of response shift. One important limitation in Paper 3 is missing data. However, this will be discussed in chapter 7.3.4.

7.3.3.1 Strengths and constraints due to the choice of sample

We chose to use data only from one of the two conditions in the RCT study. By keeping the RCT sample intact, including also the TAU condition, the power would have gone up, and it would have been possible to analyze differences between the sites. However, there are several reasons for the choice of only using data from the feedback condition. First, we only had preliminary results from the main study (Tilden et al., 2019b) at the time of analysis, so we did not know if there were any differences between the two conditions. Second, and more important is that the main interest in this study was to investigate CFT treatment where feedback was used. One of the reasons for this interest is that only the feedback sample has intersession data (i.e. data from every session) that can be used to investigate trajectories of change, something that is especially interesting for investigating cases of deterioration or no-change. However, I admit that this choice limited the possibilities to investigate the total sample only, and possible differences between the three sites were masked. Indeed, if results from all three sites were found to be similar, it would have strengthened the argument for generalization of the results.

7.3.3.2 Strengths and limitations due to the lack of a control for adherence to protocol

A limitation of this study is related to the lack of control for adherence to a specific therapeutic method proven to be efficacious, as well as the rating of competence of the therapists. The only data I have from the therapists is their theoretical orientation, where almost all identified themselves practicing family/systems therapy and that all were trained and certified according to the STIC guideline criteria. Even though this is a strength from a naturalistic standpoint, it is also a limitation because the variability of the treatment (such as therapeutic focus, specific methods, quality of alliance) may be large. However, the effect sizes calculated on the total sample in this study, showing outcomes like those of other CFT studies (Sexton et al., 2013), suggest that the delivered therapy, in general, was of satisfactory quality. Furthermore, if adherence to a

specific method were to be included in the study, I would no longer have a naturalistic study, and the results would be less generalizable. All included cases had the component of the feedback given through the STIC system. However, it is a limitation of our work in Paper 3 that we did not control for how the therapists used the feedback throughout therapy. In hindsight, I see that this lack of knowledge could have been filled by regular video recordings of the therapies.

7.3.3.3 Limitations because of the possible response shift issues

Another limitation of this study is the issues of response shift. The data used in Paper 3 was collected at two distinctly different times during treatment. The first measurement point was before therapy started and the second when the therapy was terminated. One aspect that leads to response shift is time. The other is the change in the client's lives from start to end of treatment. Response shift theory highlights that even if the wording of the question is the same at both times, the client's perspective of the matter might have changed. Hence, the change measured from start to end of therapy might as well be shift in perspective as it is a change on the matter. Indeed, change of perspective is sometimes one of the goals in therapy. One example could be a client that at start of therapy is very frustrated about the frequency of sex in their marriage. During therapy he understands the effects of his infidelity on his wife's sexual desire and also his deeper need for acceptance that lays behind his need for frequent sex. Due to this change in perspective he might not be as frustrated, even if the frequency of sex had not changed from start to end of treatment. In this way response shift could contaminate the results in Paper 3. It would have strengthened the study if "*then"* measures, as described by Howard and Dailey (1979), had been used.

7.3.4 Implications and handling of missing data

There are some missing data in the cross-sectional data used in this study, but particularly for the longitudinal data used in Paper 3, this is a topic in need of a discussion. Indeed, it is an issue often not thoroughly examined (Wood et al., 2004). I have therefore included a debate on missingness from several perspectives in this thesis. First, I discuss the simple solutions offered by the software SPSS. Second, I discuss reasons for and patterns of missingness. Third, I consider the seriousness of missingness concerning how much data that are

missing. Fourth, I discuss different methods of imputation. Finally, I discuss missing data from a philosophical perspective.

7.3.4.1 The simple solutions offered by SPSS

In the Paper 3 study, as many as 34 % of the clients answering the questionnaires at pre-treatment did not reply at post-treatment. The challenge is what to do with missing data. The simplest solution to deal with missing data is by the use of the options box of SPSS when preparing for analysis, choosing either *Replace missing data with mean, Exclude cases listwise*, or *Exclude cases pairwise*. However, the first option is suggested never to be used (Pallant, 2013), and the second option will exclude any case that has missing data on any of the variables for that case, no matter if you use a variable that is missing in the analysis or not. The listwise exclusion can thereby unnecessarily exclude cases and limit the sample size. If one of the standard options in SPSS is to be used it is the *Exclude cases pairwise* that should be chosen (Pallant, 2013). However, the issue about how to deal with missing data is complex and needs a more thorough investigation and discussion before decisions are made on how to handle them. Just to klick on the *exclude cases pairwise* option in SPSS is therefore not anything I would suggest before considering what I will discuss in the following.

7.3.4.2 Why are data missing, and do they form any pattern?

Data in this study stems from ordinary clinical practice, and some clients dropped out. Some did not finish treatment before the time of data collection was ended, and others chose not to respond to the questionnaires when they terminated treatment. I performed attrition analysis and did not find any systematic differences between the cases with complete datasets and those without, except on some minor issues. In this study, the clients were asked to complete all the questionnaires. During this procedure, when a client did not answer questions before clicking “next page,” the question/s were highlighted, and the client was asked if he/she did not want to answer. Only when clicking “yes” to that question the client could proceed to the next page. This point strengthens our assumption that the data were not missing at random. However, the seriousness of missing data depends on the pattern of the data that is missing, how much is missing and why it is missing (Tabachnick & Fidell, 2014). The pattern of the missing data is of higher importance than the amount of missing because nonrandomly missing data are serious no matter how few they are since they affect the generalizability

of the results (Tabachnick & Fidell, 2014). Therefore, Pallant (2013) argue that the researcher should consider whether the missing values were happening randomly or whether there is some systematic pattern. Missing data are normally characterized as MCAR (missing completely at random), MAR (missing at random), MNAR (missing not at random). If the data is found to be MNAR they are not to be ignored. If the missing data is predictable from the other variables the data are considered MAR. In MNAR the reason for the missingness is related to the variable itself or unobserved variables. For example, it is a well-known fact that many do not want to answer questions about their income. Thus, the missingness itself can be treated as data by creating a dummy variable. Those answering are coded as 1 and the others as 0, and then this dummy variable is used like any other variable in the analysis. However, this procedure is not necessary if the data are missing completely at random (MCAR). In Paper 3 we did run Little's test of the hypothesis that data are missing completely at random (MCAR) and got a non-significant result. We also inspected pattern of missingness on our data and were not able to detect any such patterns. The result from the Little's test and the visible inspection both suggest that treating the data as MCAR may be a reasonable assumption and that the missing data will thus not have a material impact on the accuracy of the estimates and inferences.

7.3.4.3 Does it matter how many responses that are missing?

In a big dataset, it makes sense to argue that some missing data does not matter, but there is no clear consensus on where to draw a line for when it is too much missing data (Tabachnick & Fidell, 2014). Schafer (1999) suggests 5 % missingness as a lower threshold for when multiple imputations provide negligible benefit. However, missingness at 40 % has been suggested by others (see, e.g., Jakobsen, Gluud, Wetterslev, & Winkel, 2017) to be a limit for where the missingness is too large. As mentioned, we had 34 % missing in the Paper 3 study, thus below the suggested 40 %. Madley-Dowd and colleagues (2019) argue that the level of missingness in itself is not what is most important. They exemplify in their simulation study that missingness of 1 % could be as influential as missingness at 90 % and therefore suggest assessing the fraction of missing information (FMI) instead of missingness. Their most valuable point concerning this study is that even if we have a high portion of missingness it does not automatically mean that we cannot trust the results from the analysis.

7.3.4.4 Different methods of imputing missing data

Tabachnick & Fidell (2014) suggest imputation instead of excluding missing data from the analysis, applying methods such as using prior knowledge, inserting mean values, using regression, expectation-maximization and multiple imputations. Multiple imputations are most often suggested today, and it produces several complete datasets that are being used in the analysis, and the average is then reported (Tabachnick & Fidell, 2014). The main reasons for multiple imputations to be the best option of today are because it retains sampling variability. When inspecting the data used in Paper 3 we found that the data most probably are MCAR. This point leads to the assumption that imputation would probably not bring much or any additional information, only more power to the analysis. Indeed, in our response to a reviewer on Paper 3, we did run all the analyses using multiple imputed data and got almost equivalent results as when not imputing data. However, when using Structural Equation Modeling, the missing likelihood estimation is found to be superior (Enders & Bandalos, 2001) and was the option used in Paper 1 and 2. This method does not impute any data, but rather uses each cases available data to compute estimates. The maximum likelihood estimate of a parameter is the value of the parameter that is most likely to have resulted in the observed data.

7.3.4.5 A philosophical perspective of missing data

Before I leave the subject of missing data, I want to discuss this on a philosophical level. The question to discuss is how to understand the missing data. Is a data point either missing or not missing, or could this data be viewed as a continuum? Blackwell, Honaker, and King (2012) argue that researchers should treat cell values as either observed without (random) error, observed with error, or missing. Their approach treats all values as data with an error.

They write: “measurement error is a specific type of missing data problem where observed proxy variables provide probabilistic prior information about the true unobserved cell values” (Blackwell et al., 2012, p. 2). The question is just how much uncertainty we have about any value given or not given by the clients. I.e., when a client is answering four on a Likert scale we are not sure if they mean four or if it could be 3.8 or even 4.3. In this way, uncertainty is connected to all variables in a given dataset, not only the ones missing. A missing data point could be viewed as being somewhere on the continuum, from one extreme where

we do not have any idea of what the right answer could be, to the other extreme where we are quite sure even if there are missing data in the dataset. E.g., a client has answered several different questions about how they, as a couple, tackle conflicts, but for one question about this concept, the answer is missing. We could then, based on answers to correlated questions, estimate the "true" answer to that question with good precision. According to Blackwell and colleagues (2012), we, in this case, have a (missing) data point with low uncertainty. Their solution is something they call *Multiple Overimputation*, a two-step procedure much like the Multiple Imputation method.

8 Conclusions

The main aim of this thesis was to investigate couple relationships and change in those relationships that appear during CFT treatment. This thesis addresses partly the methodological question on how such phenomena in CFT research are measured, partly how CFT change was related to the relational and societal measures, domestic violence, and work functioning.

The conclusion on the methodological question was that an extensive simplification of measures could be misleading and thus possibly challenging the use of total scores that seems to dominate this field of research when assessing change (Ogles, 2013; Sexton et al., 2013). This finding could further challenge future decisions on what focus to have when assessing change in therapy, whether individual, relational or societal.

Since this field is dominated by studies investigating individual and relational measures, I chose a societal measure to exemplify the usefulness of that focus. My choice to investigate work functioning serves as an example that might inspire researchers to investigate this and other societal measures in future projects. The highlighting of a societal measure in this thesis might also challenge CFT practitioners to attend to the client's difficulties not only at individual and relational levels but also on life functioning.

However, the part of this thesis challenging CFT practice the most may be the call to uncover the violence occurring in the clients' lives, because we found that as many as one of every four clients might suffer from physical violence. This portion, compared to registered reports, represents a considerable gap between the actual prevalence of domestic violence and what is addressed during treatment. As a prerequisite for targeting this issue properly in treatment, this phenomenon must be uncovered (Middelborg & Samoilow, 2014; Stith et al., 2011). Our finding of a connection between couple violence and the co-occurring family violence may, therefore, be helpful in clinical practice to uncover violence. More precisely, our identification of three positive predictors and one negative predictor of physical violence should work as clinical markers. A clinician can thus apply the positive predictors (expectation, anger, and sexual satisfaction) as signposts for possible physical violence occurring in the family.

Each of the three papers contributes to the main aim of this thesis, namely, to investigate couple relationships and change. Without reliable and valid measures this could not be investigated (Paper 1). Contrary, with good measures even complicated relational issues, like domestic violence can be investigated (Paper 2). The last paper (Paper 3) exemplified that when investigating something as personal as a couple relationship, a broader focus of societal importance such as work functioning is useful.

Recommendations for further research go in three directions. First, more measures should be assessed to see if the results from Paper 1 hold also stand for other ROM measures. Second, when investigating domestic violence other emotions than anger should be assessed as well as investigating the relationship between violence and satisfaction of their sexual life. Third, Paper 3 is especially calling for further research on the group of clients who do not benefit from CFT treatment but also on further exploration of the relationship between CFT and work functioning.

8.1 Final reflections

When I started this three-year-long journey I thought I would come back to the clinical staff as a more informed therapist. I now see that I am more informed but also more eager to use my newly gained knowledge of statistics and the CFT field of research for further research. I am eager to try to understand more of clients' lives and how they change in therapy, especially those who seem not to benefit from therapy. I would like to participate in the work of helping more people to a better life by practicing a combination of a therapist and a researcher.

Working with this study has also inspired me to connect to other researchers nationally and internationally. A need to facilitate for researchers within the field of CFT in Norway to meet and discuss our areas of interest and competence emerged during my study. I have therefore been involved in starting a Norwegian consortium of both quantitative and qualitative CFT researchers. The first initiative was taken by me in 2018, and the aim was to gather researchers to connect, share experiences, presenting results, learn from each other on methodology, and plan collaborative projects. To date, three meetings have been conducted, and the plan is to meet twice a year to strengthen this field of research in Norway.

9 References

Acarturk, C., Cuijpers, P., van Straten, A., & de Graaf, R. (2009). Psychological treatment of social anxiety disorder: a meta-analysis. *Psychological Medicine, 39*(2), 241-254.

Adler, D. A., McLaughlin, T. J., Rogers, W. H., Chang, H., Lapitsky, L., & Lerner, D. (2006). Job performance deficits due to depression. *American journal of psychiatry, 163*(9), 1569-1576.

Ali, P. A., Dhingra, K., & McGarry, J. (2016). A literature review of intimate partner violence and its classifications. *Aggression and violent behavior, 31*, 16-25.

Almås, E., & Pirelli Benestad, E. E. (2017). Seksualtiet som tema i parterapi. In O. Ness (Ed.), *Håndbok i parterapi* (pp. 169-181). Bergen: Fagbokforlaget.

Amnesty International. (2013). Fakta om vold mot kvinner. Retrieved from http://www.amnesty.no/aktuelt/flere-nyheter/arkiv-bakgrunn/fakta-om-vold-mot-kvinner

Andersen, A., & Svensson, T. (2013). Internet-based Mental Health Services in Norway and Sweden: Characteristics and Consequences. *Administration and Policy in Mental Health and Mental Health Services Research, 40*(2), 145-153. doi:10.1007/s10488-011-0388-2

Anderson, H. (1997). *Conversation, language, and possibilities: A postmodern approach to therapy*: Basic Books.

Anderson, H. (2019). Samarbeidende - dialogisk familieterapi. In L. Lorås & O. Ness (Eds.), *Håndbok i familieterapi* (pp. 199-212). Bergen: Fagbokforlaget.

Andrews, G. (2000). A focus on empirically supported outcomes: A commentary on search for empirically supported treatments. *Clinical Psychology: Science and Practice, 7*(3), 264-268.

Angier, N. (2007). *The canon: A whirligig tour of the beautiful basics of science*: Houghton Mifflin Company.

Anker, M. G., Duncan, B. L., Sparks, J. A., & La Greca, A. M. (2009). Using Client Feedback to Improve Couple Therapy Outcomes: A Randomized Clinical Trial in a Naturalistic Setting. *Journal of Consulting and Clinical Psychology, 77*(4), 693-704. doi:10.1037/a0016062

Anker, M. G., Owen, J., Duncan, B. L., & Sparks, J. (2010). The alliance in couple therapy: Partner influence, early change, and alliance patterns in a naturalistic sample. *J Consult Clin Psychol, 78*, 635 - 645.

Antunes-Alves, S., & de Stefano, J. (2014). Intimate partner violence: Making the case for joint couple treatment. *The Family Journal, 22*(1), 62-68.

APA-Task-Force. (2006). APA presidential task force on evidence based practice. *American Psychologist, 61*, 271-285.

Appel, A. E., & Holden, G. W. (1998). The co-occurrence of spouse and physical child abuse: A review and appraisal. *Journal of Family Psychology, 12*(4), 578.

Aronsson, G., Johansen, V., Marklund, S., Rønning, R., & Solheim, L. J. (2015). *Sjukfrånvarons dimensioner: svensk-norska jämförelser och analyser*. Stockholm: Liber.

Barnevernloven. (1993). *Lov om barneverntjenester*.

Bateson, G. (1972). *Steps to an ecology of mind*. New York: Ballantine Books.

Bateson, G., Jackson, D. D., Haley, J., & Weakland, J. (1956). Toward a theory of schizophrenia. *Behavioral science, 1*(4), 251-264.

Bauer, S., Lambert, M. J., & Nielsen, S. L. (2004). Clinical significance methods: A comparison of statistical techniques. *Journal of Personality Assessment, 82*(1), 60-70.

Beck, A. T., Steer, R., & Brown, G. (1995). *Beck Depression Inventory-II Manual*. San Antonio, TX: The Psychological Corporation.

Becker, M., Vignoles, V. L., Owe, E., Easterbrook, M. J., Brown, R., Smith, P. B., . . . Torres, A. (2018). Being oneself through time: Bases of self-continuity across 55 cultures. *Self and Identity, 17*(3), 276-293.

Bilsker, D., Gilbert, M., Myette, T. L., & Stewart-Patterson, C. (2005). *Depression & work function: bridging the gap between mental health care & the workplace.* Vancouver: Mental Health Evaluation & Community Consultation Unit at the University of British Columbia. Retrieved from https ://www.sfu.ca/carmh a/publi catio ns/depre ssion-and-work-funct ion.html

Blackwell, M., Honaker, J., & King, G. (2012). Multiple overimputation: a unified approach to measurement error and missing data. *Typescript, Harvard University*.

Blunch, N. (2013). *Introduction to structural equation modeling using IBM SPSS statistics and AMOS (second edition)*. London: Sage.

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: Access, effectiveness and efficiency: Narrative literature review. *British Journal of Psychiatry, 186*(1), 11-17. doi:10.1192/bjp.186.1.11

Bufdir, Barn ungdom og familiedepartementet. (2017, 04.05.2017). Om partnervold. Retrieved from https://www.bufdir.no/vold/Vold_og_overgrep_mot_voksne/Partnervold/

Busby, D. M., Christensen, C., Crane, D. R., & Larson, J. H. (1995). A revision of the dyadic adjustment scale for use with distressed and nondistressed couples: construct hierarchy and multidimensional scales. *Journal of Marital and Family Therapy, 21*(3), 289-308. doi:10.1111/j.1752-0606.1995.tb00163.x

Butzer, B., & Campbell, L. (2008). Adult attachment, sexual satisfaction, and relationship satisfaction: A study of married couples. *Personal relationships, 15*(1), 141-154.

Campbell, A. (2003). Some thoughts on methodologies and practice research networks. *Australian and New Zealand Journal of Family Therapy, 24*(1), 51-52.

Cano, A., Avery-Leaf, S., Cascardi, M., & O'Leary, K. D. (1998). Dating violence in two high school samples: Discriminating variables. *Journal of Primary Prevention, 18*(4), 431-446.

Capaldi, D. M., Knoble, N. B., Shortt, J. W., & Kim, H. K. (2012). A systematic review of risk factors for intimate partner violence. *Partner abuse, 3*(2), 231-280.

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

Carr, A. (2014b). The evidence base for family therapy and systemic interventions for child-focused problems. *Journal of Family Therapy, 36*(2), 107-157. doi:10.1111/1467-6427.12032

Carr, A. (2019). Forskning på familieterapi. In L. Lorås & O. Ness (Eds.), *Håndbok i familieterapi* (pp. 105-117). Bergen: Fagbokforlaget.

Carr, A., & Stratton, P. (2017). The SCORE family assessment questionnaire: a decade of progress. *Family process, 56*(2), 285-301.

Cascardi, M., Langhinrichsen, J., & Vivian, D. (1992). Marital aggression: Impact, injury, and health correlates for husbands and wives. *Archives of internal medicine, 152*(6), 1178-1184.

Clarkin, J. F., & Levy, K. N. (2004). The influence of client variables on psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfields Handbook of Psychotherapy and Behavior Change, Fifth edition* (Fifth edition ed., pp. 854). New York: John Wiley & Sons.

Coker, A. L., Davis, K. E., Arias, I., Desai, S., Sanderson, M., Brandt, H. M., & Smith, P. H. (2002). Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine, 23*(4), 260-268.

Coker, A. L., Smith, P. H., McKeown, R. E., & King, M. J. (2000). Frequency and correlates of intimate partner violence by type: physical, sexual, and psychological battering. *American journal of public health, 90*(4), 553.

Collier, A. (1994). Critical realism: an introduction to Roy Bhaskar's philosophy.

Comer, J. S., & Kendall, P. C. (2004). A symptom-level examination of parent–child agreement in the diagnosis of anxious youths. *Journal of the American Academy of Child & Adolescent Psychiatry, 43*(7), 878-886.

Comer, J. S., & Kendall, P. C. (2013). Methodology, design, and evalutation in psychotherapy research. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy research and behavior change, 6th ed.* (6th edition ed., pp. 21-48). New Jersey: Wiley.

Crane, D. R. (2011). Does family therapy reduce health care costs for more than the identified patient? *Clinical child psychology and psychiatry, 16*(1), 3-4.

Crane, D. R., & Christenson, J. D. (2012). A summary report of the cost-effectiveness of the profession and practice of marriage and family therapy. *Contemporary Family Therapy, 34*(2), 204-216.

Crane, D. R., & Payne, S. H. (2011). Individual versus family psychotherapy in managed care: Comparing the costs of treatment by the mental health professions. *Journal of Marital and Family Therapy, 37*(3), 273-289.

Cuijpers, P., Sijbrandij, M., Koole, S., Huibers, M., Berking, M., & Andersson, G. (2014). Psychological treatment of generalized anxiety disorder: a meta-analysis. *Clinical Psychology Review, 34*(2), 130-140.

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychological Medicine, 40*(2), 211-223.

de Flon, H. (2019). Hva er familieterapi? In L. Lorås & O. Ness (Eds.), *Håndbok i familieterapi* (pp. 15-24). Bergen: Fagbokforlaget.

DeMaris, A., Benson, M. L., Fox, G. L., Hill, T., & Van Wyk, J. (2003). Distal and proximal factors in domestic violence: A test of an integrated model. *Journal of Marriage and Family, 65*(3), 652-667.

Desmarais, S. L., Reeves, K. A., Nicholls, T. L., Telford, R. P., & Fiebert, M. S. (2012a). Prevalence of physical violence in intimate relationships, Part 1: Rates of male and female victimization. *Partner abuse, 3*(2), 140-169.

Desmarais, S. L., Reeves, K. A., Nicholls, T. L., Telford, R. P., & Fiebert, M. S. (2012b). Prevalence of physical violence in intimate relationships, Part 2: Rates of male and female perpetration. *Partner abuse, 3*(2), 170-198.

DeWall, C. N., Baumeister, R. F., Stillman, T. F., & Gailliot, M. T. (2007). Violence restrained: Effects of self-regulation and its depletion on aggression. *Journal of Experimental social psychology, 43*(1), 62-76.

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and perspectives, 38*(1), 105.

Dube, S. R., Anda, R. F., Felitti, V. J., Croft, J. B., Edwards, V. J., & Giles, W. H. (2001). Growing up with parental alcohol abuse: exposure to childhood abuse, neglect, and household dysfunction. *Child abuse & neglect, 25*(12), 1627-1640.

Duncan, B. L., Miller, S., Sparks, J., Claud, D., Reynolds, L., & Brown, J. (2003). The Session Rating Scale: Preliminary psychometric properties of a 'working' alliance measure. *J Brief Ther, 3*, 3 - 12.

Duncan, B. L., Miller, S. D., & Sparks, J. (2004). *The heroic client : a revolutionary way to improve effectiveness through client-directed, outcome-informed therapy* (Rev. ed. ed.). San Francisco: Jossey-Bass.

Duncan, B. L., & Reese, R. J. (2013). Empirically Supported Treatments, Evidence- Based Treatments, and Evidence- Based Practice. In W. I. B. (Ed.), *Handbook of Psychology, Second Edition* (Vol. 8, pp. 489-513). Hoboken. NJ. USA: John Wiely & Sons, Inc.

Ehrensaft, M. K., & Vivian, D. (1996). Spouses' reasons for not reporting existing marital aggression as a marital problem. *Journal of Family Psychology, 10*(4), 443.

Eisler, I., Dare, C., Hodes, M., Russell, G., Dodge, E., & Le Grange, D. (2000). Family therapy for adolescent anorexia nervosa: the results of a controlled comparison of two family interventions. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(6), 727-736.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling, 8*(3), 430-457.
Entilli, L., & Cipolletta, S. (2017). When the woman gets violent: the construction of domestic abuse experience from heterosexual men's perspective. *Journal of clinical nursing, 26*(15-16), 2328-2341.
Epstein, N. B., Baldwin, L. M., & Bishop, D. S. (1983). The McMaster family assessment device. *Journal of Marital and Family Therapy, 9*(2), 171-180.
Eurostat. (2015) Marriage and divorce statistics. Retrieved from: https://ec.europa.eu/eurostat/statistics-explained/index.php/Marriage_and_divorce_statistics
Fals-Stewart, W. (2003). The occurrence of partner physical aggression on days of alcohol consumption: A longitudinal diary study. *Journal of Consulting and Clinical Psychology, 71*(1), 41.
Fals-Stewart, W., Kashdan, T. B., O'Farrell, T. J., & Birchler, G. R. (2002). Behavioral couples therapy for drug-abusing patients: Effects on partner violence. *Journal of substance abuse treatment, 22*(2), 87-96.
Fals-Stewart, W., O'Farrell, T. J., Birchler, G. R., Córdova, J., & Kelley, M. L. (2005). Behavioral couples therapy for alcoholism and drug abuse: Where we've been, where we are, and where we're going. *Journal of Cognitive Psychotherapy, 19*(3), 229-246.
Felder, R. M., & Spurlin, J. (2005). Applications, reliability and validity of the index of learning styles. *International journal of engineering education, 21*(1), 103-112.
Field, A. (2018). *Discovering statistics using SPSS, 5th edition* (5th ed.). London: Sage publications.
Fusco, R. A., & Fantuzzo, J. W. (2009). Domestic violence crimes and children: A population-based investigation of direct sensory exposure and the nature of involvement. *Children and Youth Services Review, 31*(2), 249-256.
Gausel, N. (2013). Self-reform or self-defense? Understanding how people cope with their moral failures by understanding how they appraise and feel about their moral failures. In M. Moshe & N. Corbu (Eds.), *Walk of shame* (pp. 191-208). NY, USA: Hauppauge.
Gausel, N., & Leach, C. W. (2011). Concern for self- image and social image in the management of moral failure: Rethinking shame. *European Journal of Social Psychology, 481*(4), 62-72. doi:10.1002/ejsp.803
Gausel, N., Leach, C. W., Mazziotta, A., & Feuchte, F. (2018). Seeking revenge or seeking reconciliation? How concern for social- image and felt shame helps explain responses in reciprocal intergroup conflict. *European Journal of Social Psychology, 48*(1), O62-O72.
Gergen, K. J. (2015). *An invitation to social construction*. Los Angeles: SAGE.
Golding, J. M. (1999). Intimate partner violence as a risk factor for mental disorders: A meta-analysis. *Journal of Family Violence, 14*(2), 99-132.

Gottman, J. M., & Silver, N. (2015). *The seven principles for making marriage work: A practical guide from the country's foremost relationship expert*: Harmony.
Gottman, J. S., & Gottman, J. M. (2015). *10 Principles for Doing Effective Couples Therapy*. New York: WW Norton & Company.
Gullestad, S. (2001). Hva er evidensbasert psykoterapi? *Tidsskrift for Norsk psykologforening, 38*(10), 942-951.
Gurman, A. S. (2008). A framework for the comparative study of couple therapy. In A. S. Gurman (Ed.), *Clinical handbook of couple therapy* (pp. 1-26). New York: Guilford Press.
Gurman, A. S. (2015). The theory and practice of couple therapy. In A. S. Gurman (Ed.), *Clinical handbook of couple therapy (5. utg.)* (5 ed., pp. 1-18). New York: The Guilford Press.
Gurman, A. S., Kniskern, D., & Pinsof, W. (1986). Research on the process and outcome of marital and family therapy. In S. Garfield & A. E. Bergin (Eds.), *Psychotherapy and behaviour change* (3 ed., pp. 565-625). New York: John Wilwy & Sons.
Gurman, A. S., Lebow, J. L., & Snyder, D. K. (2015). *Clinical handbook of couple therapy*. New York: Guilford Publications.
Haaland, T., Clausen, S.-E., & Schei, B. (2005). *Vold i parforhold - ulike perspektiver : resultater fra den første landsdekkende undersøkelsen i Norge* (Vol. 2005:3). Oslo: Norsk institutt for by- og regionforskning.
Helgeland, A. (2014). *Profesjonsetiske utfordringer ved barns deltakelse i familieterapi*. Kristiansand: University of Agder.
Helsedirektoratet. (2015a). Prioriteringsveileder - Psykisk helsevern for barn og unge. Retrieved from: https://www.helsedirektoratet.no/veiledere/prioriteringsveiledere/psykisk-helsevern-for-barn-og-unge
Helsedirektoratet. (2015b). Prioriteringsveileder - Psykisk helsevern for voksne. Retrieved from: https://www.helsedirektoratet.no/veiledere/prioriteringsveiledere/psykisk-helsevern-for-voksne
Henderson, M., Harvey, S. B., Øverland, S., Mykletun, A., & Hotopf, M. (2011). Work and common psychiatric disorders. *Journal of the Royal Society of Medicine, 104*(5), 198-207. doi:10.1258/jrsm.2011.100231
Hensing, G., Andersson, L., & Brage, S. (2006). Increase in sickness absence with psychiatric diagnosis in Norway: a general population-based epidemiologic study of age, gender and regional distribution. *Bmc Medicine, 4*(1), 19.
Hiller, W., Schindler, A. C., & Lambert, M. J. (2012). Defining response and remission in psychotherapy research: a comparison of the RCI and the method of percent improvement. *Psychotherapy Research, 22*(1), 1-11.
Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology, 64*(2), 144.

Husso, M., Virkki, T., Notko, M., Holma, J., Laitila, A., & Mäntysaari, M. (2012). Making sense of domestic violence intervention in professional health care. *Health & social care in the community, 20*(4), 347-355.
Hårtveit, H., & Jensen, P. (2004). *Familien - pluss én : innføring i familieterapi* (2. utg. ed.). Oslo: Universitetsforl.
Jacobson, N. S., & Truax, P. (1991). Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19.
Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology, 17*(1), 162.
Jenkinson, C., Stewart-Brown, S., Petersen, S., & Paice, C. (1999). Assessment of the SF-36 version 2 in the United Kingdom. *Journal of Epidemiology & Community Health, 53*(1), 46-50.
Jensen, P. (2006). Hvordan kan «den store psykoterapidebatten» påvirke forståelsen av familieterapiforskningen? Hvilken terapeut har det beste tilbudet til hvilken klient?
Jewkes, R. (2002). Intimate partner violence: causes and prevention. *The Lancet, 359*(9315), 1423-1429.
Johnsen, A., & Torsteinsson, V. W. (2012). *Lærebok i familieterapi*. Oslo: Universitetsforl.
Johnson, M. P., & Ferraro, K. J. (2000). Research on domestic violence in the 1990s: Making distinctions. *Journal of Marriage and Family, 62*(4), 948-963.
Jose, A., & O'Leary, K. D. (2009). Prevalence of partner aggression in representative and clinic samples.
Jouriles, E. N., & O'Leary, K. D. (1985). Interspousal reliability of reports of marital violence. *Journal of Consulting and Clinical Psychology, 53*(3), 419.
Justis- og beredskapsdepartementet. (2014). Et liv uten vold: Handlingsplan mot vold i nære relasjoner 2014-2017.
Karakurt, G., Whiting, K., Van Esch, C., Bolen, S. D., & Calabrese, J. R. (2016). Couples therapy for intimate partner violence: A systematic review and meta- analysis. *Journal of Marital and Family Therapy, 42*(4), 567-583.
Kazdin, A. E. (2003). Psychotherapy for children and adolescents. *Annual Review of Psychology, 54*(1), 253-276.
Kazdin, A. E. (2008). Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63*(3), 146.
Kendrick, T., El- Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., . . . de Jong, K. (2016). Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database of Systematic Reviews*(7).

Kimball, E. (2016). Edleson Revisited: Reviewing Children's Witnessing of Domestic Violence 15 Years Later. *Journal of Family Violence, 31*(5), 625-637. doi:10.1007/s10896-015-9786-7

Kitzmann, K. M., Gaylord, N. K., Holt, A. R., & Kenny, E. D. (2003). Child witnesses to domestic violence. A meta-analytic review. *Journal of Consulting and Clinical Psychology, 71*(2), 339-352.

Klein, J. B., Jacobs, R. H., & Reinecke, M. A. (2007). Cognitive-behavioral therapy for adolescent depression: a meta-analytic investigation of changes in effect-size estimates. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*(11), 1403-1413.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford publications.

Kraus, D. R., Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of Clinical Psychology, 61*(3), 285-314.

Krisesentersekretariatet. (2017). Statistikk. Retrieved from http://www.krisesenter.com/backup240616/tall-og-fakta/nasjonal-statistikk

Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). *World report on violence and health*. Geneva: World Health Organization.

Lambert, M. J. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research, 17*, 1-14.

Lambert, M. J. (2010). Yes, it is time for clinicians to routinely monitor treatment outcome.

Lambert, M. J. (2013a). *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th. ed.): John Wiley & Sons.

Lambert, M. J. (2013b). Introduction and historical overview. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (sicth edition ed., pp. 3-20). New Jersey: John Wiley & Sons Inc.

Lambert, M. J. (2015). Progress feedback and the OQ-system: The past and the future. *Psychotherapy, 52*(4), 381.

Lambert, M. J., Burlingame, G., Umphress, V., Hansen, N., Vermeersch, D., & Clouse, G. (1996). The reliability and validity of the Outcome Questionnaire. *Clin Psychol Psychother, 3*, 249 - 258.

Lambert, M. J., & Finch, A. E. (1999). The Outcome Questionnaire.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69*(2), 159.

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy, 48*, 72 - 79.

Lambert, M. J., Whipple, J., Smart, D., Vermeersch, D., & Nielsen, S. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychother Res, 11*, 49 - 68.

Langkaas, T. F., Wampold, B. E., & Hoffart, A. (2018). Five types of clinical difference to monitor in practice. *Psychotherapy, 55*(3), 241.
Larner, G. (2004). Family therapy and the politics of evidence. *Journal of Family Therapy, 26*(1), 17-39.
Leary, M. R., Twenge, J. M., & Quinlivan, E. (2006). Interpersonal rejection as a determinant of anger and aggression. *Personality and Social Psychology Review*(10(2)), 111-132.
Lerner, D., & Henke, R. M. (2008). What Does Research Tell Us About Depression, Job Performance, and Work Productivity? *Journal of Occupational and Environmental Medicine, 50*(4), 401-410. doi:10.1097/JOM.0b013e31816bae50
Loge, J. H., & Kaasa, S. (1998). Short form 36 (SF-36) health survey: normative data from the general Norwegian population. *Scandinavian journal of social medicine, 26*(4), 250-258.
Loras, L. (2016). *The Map of Competencies in Systemic Therapy: A Qualitative Study of the Systemic Competencies in Norwegian Child and Adolescent Mental Health that Target the Associated Abnormal Psychosocial Situations in Axis 5 (ICD-10)*. London: University of East London.
Lorås, L., & Ness, O. (2019). *Håndbok i familieterapi*. Bergen: Fagbokforlaget.
Maddigan, S., & Nylund, D. (2019). Narrativ familieterapi. In L. Lorås & O. Ness (Eds.), *Håndbok i familieterapi* (pp. 183-197). Bergen: Fagbokforlaget.
Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of clinical epidemiology*. doi:https://doi.org/10.1016/j.jclinepi.2019.02.016
Mazziotta, A., Feuchte, F., Gausel, N., & Nadler, A. (2014). Does remembering past ingroup harmdoing promote postwar cross- group contact? Insights from a field- experiment in Liberia. *European Journal of Social Psychology, 44*(1), 43-52.
McLeod, J. (2013). Qualitative research: Methods and contributions. In M. J. Lambert (Ed.), *Handbook of psychotherapy and behaviour change* (6 ed., pp. 49-84). New Jersey: John Wiley & Sons.
Mederos, F. (2002). Changing our visions of intervention: The evolution of programs for physically abusive men. In E. Aldarondo & F. Mederos (Eds.), *Programs for men who batter: Intervention and prevention strategies in a diverse society* (pp. 1-23). Kingston, NJ: Civic research institute.
Middelborg, J., & Samoilow, D. K. (2014). *Tryggere barndom : et behandlingsperspektiv på vold i familien*. Oslo: Gyldendal akademisk.
Miller, S., Duncan, B. L., Brown, J., Sorrell, R., & Chalk, M. (2006). Using formal client feedback to improve retention and outcome: Making ongoing, real-time assessment feasible. *J Brief Ther, 5*, 5 - 22.
Miller, S., Duncan, B. L., Brown, J., Sparks, J., & Claud, D. (2003). The outcome rating scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *J Brief Ther, 2*, 91 - 100.

Morfeld, M., Bullinger, M., Nantke, J., & Brähler, E. (2005). The version 2.0 of the SF-36 Health Survey: results of a population-representative study. *Sozial-und Praventivmedizin, 50*(5), 292-300.

Morgan, H., & Russell, G. (1975). Value of family background and clinical features as predictors of long-term outcome in anorexia nervosa: four-year follow-up study of 41 patients. *Psychological Medicine, 5*(4), 355-371.

Mossige, S., & Stefansen, K. (2007). *Vold og overgrep mot barn og unge : en selvrapporteringsstudie blant avgangselever i videregående skole* NOVA-rapport (online), Vol. 20/2007.

Ness, O. (2011). *Learning new ideas and practices together: A co-operative inquiry into learning to use Johnella Bird's relational language-making approach in couples therapy*. Tilburg: Tilburg University Press.

Ness, O. (2017). *Håndbok i parterapi* (O. Ness Ed.). Bergen: Fagbokforlaget.

NKPH, N. k. f. p. h. (2014). Vold og overgrep i nære relasjoner. Retrieved from http://www.napha.no/content/13928/Vold-og-overgrep-i-nare-relasjoner

Novak, J. R., Sandberg, J. G., & Davis, S. Y. (2017). The Role of Attachment Behaviors in the Link between Relationship Satisfaction and Depression in Clinical Couples: Implications for Clinical Practice. *Journal of Marital and Family Therapy, 43*(2), 352-363. doi:10.1111/jmft.12201

O'Farrell, T. J., Murphy, C. M., Stephan, S. H., Fals-Stewart, W., & Murphy, M. (2004). Partner violence before and after couples-based alcoholism treatment for male alcoholic patients: the role of treatment involvement and abstinence. *Journal of Consulting and Clinical Psychology, 72*(2), 202.

O'Hanrahan, K., Daly White, M., Carr, A., Cahill, P., Keenleyside, M., Fitzhenry, M., . . . O'Shea, H. (2017). Validation of 28 and 15 item versions of the SCORE family assessment questionnaire with adult mental health service users. *Journal of Family Therapy, 39*(1), 4-20.

O'Leary, K. D. (2001). Conjoint therapy for partners who engage in physically aggressive behavior: Rationale and research. *Journal of Aggression, Maltreatment & Trauma, 5*(2), 145-164.

O'Leary, K. D., Vivian, D., & Malone, J. (1992). Assessment of physical aggression against women in marriage: The need for multimodal assessment. *Behavioral Assessment*.

O'Leary, K. D., & Woodin, E. M. (2009). *Psychological and physical aggression in couples: Causes and interventions*: American Psychological Association.

Oanes, C. J. (2016). *Tilbakemelding i terapi. På hvilke måter opplever terapeuter at tilbakemeldingsprosedyrer kan virke inn på terapeutiske prosesser?* Bergen: University of Bergen.

OECD. (2003). *Transforming disability into ability. Policies to promote work and income security for disabled people.* Paris: OECD Publications Service.

Ogles, B. M. (2013). Measuring change in psychotherapy research. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of psychotherapy research and behavior change 6th ed.* (pp. 134-166). New Jersey: Wiley.

Ogles, B. M., Lunnen, K. M., & Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review, 21*(3), 421-446.
Olkowska, A., Sundet, R., & Karlsson, B. E. (2018). Kan klient- og resultatstyrt praksis (KOR) bidra til økt brukermedvirkning og med dette til bedrings- og mestringsprosesser i terapi? *Fokus på familien, 46*(04), 294-315. doi:10.18261/issn.0807-7487-2018-04-05
Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of life research, 14*(3), 587-598.
Organization, W. H. (2002). *The world health report 2002: reducing risks, promoting healthy life*: World Health Organization.
Ormhaug, S. M., Jensen, T. K., Hukkelberg, S. S., & Egeland, K. (2012). Traumer hos barn - blir de gjemt eller glemt? ; kartlegging av traumatiske erfaringer hos barn og unge henvist til BUP. *Tidsskrift for Norsk psykologforening, 49*(3), 235-240.
Owe, E., Vignoles, V. L., Becker, M., Brown, R., Smith, P. B., Lee, S. W., . . . Gheorghiu, M. (2013). Contextualism as an important facet of individualism-collectivism: Personhood beliefs across 37 national groups. *Journal of Cross-Cultural Psychology, 44*(1), 24-45.
Owen, J., Duncan, B. L., Reese, R. J., Anker, M., & Sparks, J. (2014). Accounting for therapist variability in couple therapy outcomes: What really matters? *Journal of Sex & Marital Therapy, 40*(6), 488-502.
Pallant, J. (2013). *SPSS survival manual : a step by step guide to data analysis using IBM SPSS* (5th ed. ed.). Maidenhead: McGraw-Hill.
Pan, H. S., Neidig, P. H., & O'leary, K. D. (1994). Predicting mild and severe husband-to-wife physical aggression. *Journal of Consulting and Clinical Psychology, 62*(5), 975.
Pape, H., & Hjemdal, O. K. (2004). *Den skjulte volden?* Vol. 1/2004. *En undersøkelse av Oslobefolkningens utsatthet for trusler, vold og seksuelle overgrep*
Pinsof, W. M. (1992). Toward a scientific paradigm for family psychology: The integrative process systems perspective. *Journal of Family Psychology, 5*(3-4), 432.
Pinsof, W. M. (2017). The Systemic Therapy Inventory of Change—STIC: A Multi-systemic and Multi-dimensional System to Integrate Science into Psychotherapeutic Practice *Routine Outcome Monitoring in Couple and Family Therapy* (pp. 85-101): Springer.
Pinsof, W. M., Breunlin, D. C., Chambers, A. L., Solomon, A. H., & Russell, W. P. (2015a). Integrative Problem-Centered Metaframeworks Approach. In A. S. Gurman, J. Lebow, & D. K. Snyder (Eds.), *Clinical Handbook of Couple Therapy* (pp. 161-191). New York: The Guilford Press.
Pinsof, W. M., & Lebow, J. L. (2005). A Scientific Paradigm for Family Psychology. In W. M. Pinsof & J. L. Lebow (Eds.), *The art of the science* (pp. 3-19). New York: Oxford University Press.

Pinsof, W. M., & Wynne, L. C. (2000). Toward progress research: Closing the gap between family therapy practice and research. *Journal of Marital and Family Therapy, 26*(1), 1-8.

Pinsof, W. M., Zinbarg, R. E., Lebow, J., Knobloch-Fedders, L. M., Durbin, E., Chambers, A., . . . Friedman, G. (2009). Laying the foundation for progress research in family, couple, and individual therapy: The development and psychometric features of the initial Systemic Therapy Inventory of Change. *Psychotherapy Research, 19*(2), 143-156.

Pinsof, W. M., Zinbarg, R. E., Shimokawa, K., Latta, T. A., Goldsmith, J. Z., Knobloch-Fedders, L. M., . . . Lebow, J. L. (2015b). Confirming, Validating, and Norming the Factor Structure of Systemic Therapy Inventory of Change Initial and Intersession. *Family process, 54*(3), 464-484. doi:10.1111/famp.12159

Powers, M. B., Vedel, E., & Emmelkamp, P. M. (2008). Behavioral couples therapy (BCT) for alcohol and drug use disorders: A meta-analysis. *Clinical Psychology Review, 28*(6), 952-962.

Punch, K. F. (2014). *Introduction to social research: Quantitative and qualitative approaches*. London: Sage.

Raundalen, M. (2009). Vold mot barn - skader. *Bulleteng nr. 6,* pp. 2-32.

Reese, R., Toland, M., Slone, N., & Norsworthy, L. (2010). Effect of client feedback on couple psychotherapy outcomes. *Psychother Theory Res Pract Train, 47*, 616 - 630.

Reigstad, B., Jørgensen, K., & Wichstrøm, L. (2006). Diagnosed and self-reported childhood abuse in national and regional samples of child and adolescent psychiatric patients: Prevalences and correlates. *Nordic Journal of Psychiatry, 60*(1), 58-66. doi:10.1080/08039480500504933

Renner, L. M., & Slack, K. S. (2006). Intimate partner violence and child maltreatment: Understanding intra-and intergenerational connections. *Child abuse & neglect, 30*(6), 599-617.

Rosenbaum, A., & Kunkel, T. S. (2009). Group intervetions for intimate partner violence. In K. O'Leary & E. M. Woodin (Eds.), *Psychological and physical aggression in coules. Causes and interventions* (pp. 191-210). Washington, DC: American psychological association.

Rotunda, R. J., Scherer, D. G., Imm, P. S., & Deleon, P. H. (1995). Family Systems and Alcohol Misuse: Research on the Effects of Alcoholism on Family Functioning and Effective Family Interventions. *Professional Psychology: Research and Practice, 26*(1), 95-104. doi:10.1037/0735-7028.26.1.95

Russell, G. F., Szmukler, G. I., Dare, C., & Eisler, I. (1987). An evaluation of family therapy in anorexia nervosa and bulimia nervosa. *Archives of General Psychiatry, 44*(12), 1047-1056.

Sandmo, E. (1999). *Voldssamfunnets undergang: om disiplineringen av Norge på 1600-tallet*: Universitetsforl.

Sandvig, J. (2009, 23.12.09). Kommune dømt for passivitet i barnevernssak. *Aftenposten*. Retrieved from

https://www.aftenposten.no/norge/i/2d7AG/Kommune-domt-for-passivitet-i-barnevernssak
Schafer, J. L. (1999). Multiple imputation: a primer. *8*(1), 3-15. doi:10.1177/096228029900800102
Schiepek, G., Eckert, H., & Weihrauch, S. (2003). Critical Fluctuations and Clinical Change: Data-Based Assessment in Dynamic Systems1. *Constructivism in the Human Sciences, 8*(1), 57.
Schiepek, G. K., Tominschek, I., & Heinzel, S. (2014). Self-organization in psychotherapy: testing the synergetic model of change processes. *Frontiers in psychology, 5*, 1089.
Schumm, J. A., O'Farrell, T. J., Murphy, C. M., & Fals-Stewart, W. (2009). Partner violence before and after couples-based alcoholism treatment for female alcoholic patients. *Journal of Consulting and Clinical Psychology, 77*(6), 1136.
Schwartz, C. E. (2010). Applications of response shift theory and methods to participation measurement: a brief history of a young field. *Archives of Physical Medicine and Rehabilitation, 91*(9), S38-S43.
Schwartz, C. E., Andresen, E. M., Nosek, M. A., Krahn, G. L., & Measurement, R. E. P. o. H. S. (2007). Response shift theory: important implications for measuring quality of life in people with disability. *Archives of Physical Medicine and Rehabilitation, 88*(4), 529-536.
Schwartz, C. E., Bode, R., Repucci, N., Becker, J., Sprangers, M. A., & Fayers, P. M. (2006). The clinical significance of adaptation to changing health: a meta-analysis of response shift. *Quality of life research, 15*(9), 1533-1550.
Sellers, C. S. (1999). Self- control and intimate violence: An examination of the scope and specification of the general theory of crime. *Criminology, 37*(2), 375-404.
Senn, S. (2018). Statistical pitfalls of personalized medicine: Nature Publishing Group.
Setzer, M., & Rürup, B. (2013). Consequences for Unemployment. *Facetten der Globalisierung: Ökonomische, soziale und politische Aspekte*, 119.
Sexton, T. L., Alexander, J., & Mease, A. L. (2004). Levels of evidence for the models and mechanisms of therapeutic change in family and couple therapy. *Bergin and Garfield's handbook of psychotherapy and behavior change, 5*, 590-646.
Sexton, T. L., Datchi, C., Evans, L., LaFollette, J., & Wright, L. (2013). The effectiveness of couple and family-based clinical interventions. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of psychotherapy and behavior change, 6th edition.* (6th edition ed., pp. 587-639). New Jersey: John Wiley & Sons.
Shadish, W. R., & Baldwin, S. A. (2002). Meta-analysis of marriage and family therapy. In D. H. Sprenkle (ed.), Effectiveness research in marriage and family therapy (pp. 339-370). Alexandria, VA: The American Association for Marriage and Family Therapy.

Simpson, L. E., Atkins, D. C., Gattis, K. S., & Christensen, A. (2008). Low-level relationship aggression and couple therapy outcomes. *Journal of Family Psychology, 22*(1), 102-111.
Simpson, L. E., Doss, B. D., Wheeler, J., & Christensen, A. (2007). Relationship violence among couples seeking therapy: Common couple violence or battering? *Journal of Marital and Family Therapy, 33*(2), 270-283.
Slep, A. M. S., & O'leary, S. G. (2005). Parent and partner violence in families with young children: rates, patterns, and connections. *Journal of Consulting and Clinical Psychology, 73*(3), 435.
Solem, M.-B. (2002). Å ta empirien med en klype salt: Sentrale problemer i familieforskning i barne og ungdomspsykiatrien. *Fokus på familien, 31*(01), 34-45.
Solem, M.-B., Thuen, F., & Tilden, T. (2008). Å ta empirien på alvor – Er forskningsbasert kunnskap relevant for systemiske familiebehandlere? *Fokus på familien, 36*(2), 88-106.
Sparks, J. A. (2015). The Norway couple project: Lessons learned. *Journal of Marital and Family Therapy, 41*(4), 481-494.
Sprecher, S. (2002). Sexual satisfaction in premarital relationships: Associations with satisfaction, love, commitment, and stability. *Journal of sex research, 39*(3), 190-196.
SSB, S. S. (2016). Familievern, 2014. Retrieved from http://www.ssb.no/sosiale-forhold-og-kriminalitet/statistikker/famvern/ aar/2015-10-27?fane=tabell&sort=nummer&tabell=243787
Staub, E. (2005). The Origins and Evolution of Hate, With Notes on Prevention. In R. J. Sternberg (Ed.), *The psychology of hate* (pp. 51-66). Washington, DC, US: American Psychological Association.
Stith, S. M., & McCollum, E. E. (2009). Couples treatment for psychological and physical aggression *Psychological and physical aggression in couples: Causes and interventions* (pp. 233-250). Washington, DC: American Psychological Association; US.
Stith, S. M., McCollum, E. E., Amanor- Boadu, Y., & Smith, D. (2012). Systemic perspectives on intimate partner violence treatment. *Journal of Marital and Family Therapy, 38*(1), 220-240.
Stith, S. M., McCollum, E. E., & Rosen, K. H. (2011). *Couples therapy for domestic violence: Finding safe solutions*: American Psychological Association.
Stratton, P., Lask, J., Bland, J., Nowotny, E., Evans, C., Singh, R., . . . Peppiatt, A. (2014). Detecting therapeutic improvement early in therapy: validation of the SCORE- 15 index of family functioning and change. *Journal of Family Therapy, 36*(1), 3-19.
Sundet, R. (2009). *Client directed, outcome informed therapy in an intensive family therapy unit - A study of the use of research generated knowledge in clinical practice*. PhD thesis. Oslo: University of Oslo.
Sundet, R. (2019). I den evidensbaserte bobla. *Tidsskrift for Norsk psykologforening, 56*(9), 680-682.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using Multivariate Statistics, 6th edt*. Harlow: Pearson Education Limited.

Takahashi, M., Tanaka, K., & Miyaoka, H. (2006). Reliability and validity of communication skills questionnaire (CSQ). *Psychiatry and Clinical Neurosciences, 60*(2), 211-218.

Tasma, M., Swart, M., Wolters, G., Liemburg, E., Bruggeman, R., Knegtering, H., & Castelein, S. (2016). Do routine outcome monitoring results translate to clinical practice? A cross-sectional study in patients with a psychotic disorder. *BMC Psychiatry, 16*(1), 107.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. London: Sage.

Thagaard, T. (2003). *Systematikk og innlevelse - en innføring i kvalitativ metode (2. utgave)* (2 ed.). Bergen: Fagbokforlaget Vigmostad og Bjørke AS.

Thoresen, S., & Hjemdal, O. K. (2014). *Vold og voldtekt i Norge* Vol. 1. *En nasjonal forekomststudie av vold i et livsløpsperspektiv*

Tilden, T. (2010). The course and outcome of dyadic adjustment and individual distress during and after residential couple therapy.

Tilden, T. (2017). How do I know whether my efforts are helpful for the client? Implementing feedback in Norway. In T. Tilden & B. Wampold (Eds.), *Routine ouitcome monitoring in couple and family therapy: The empirically informed therapist* (pp. 3-13). Cham: Springer.

Tilden, T. (2019). Kognitiv familieterapi. In L. Lorås & O. Ness (Eds.), *Håndbok i familieterapi* (pp. 369-381). Bergen: Fagbokforlaget.

Tilden, T., Håland, Å. T., Hunnes, K., Fossli, G., & Oanes, C. J. (2015). Utprøving av systematisk tilbakemelding i par- og familieterapi: Barrierer og utfordringer. *Fokus på familien, 43*, 292-312.

Tilden, T., Håland, Å. T., Hunnes, K., Lærum, K., Finset, A., & Gude, T. (2010). Internett-basert tilbakemelding i par- og familieterapi ; erfaringer med empiriinformert terapi. *Tidsskrift for Norsk psykologforening*.

Tilden, T., Theisen, M., Wampold, B. E., Johnson, S. U., & Hoffart, A. (2019a). Individual distress and dyadic adjustment over the course of couple therapy and three-year follow-up: A replication study. *Psychotherapy Research*, 1-12.

Tilden, T., & Wampold, B. E. (2017). *Routine Outcome Monitoring in Couple and Family Therapy: The Empirically Informed Therapist*: Springer.

Tilden, T., Wampold, B. E., Ulvenes, P. G., Zahl-Olsen, R., Hoffart, A., Barstad, B., . . . Håland, Å. T. (2019b). Feedback in Couple and Family Therapy: A Randomized Clinical Trial. *Family process*. doi:10.1111/famp.12485

Trust, M. O. (2016). What is the SF-36v2. Retrieved from http://www.sf-36.org/tools/

Ulvestad, A. K. (2007). *Klienten - den glemte terapeut : brukerstyring i psykisk helsearbeid*. Oslo: Gyldendal akademisk.

Vall, B., Seikkula, J., Laitila, A., & Holma, J. (2016). Dominance and dialogue in couple therapy for psychological intimate partner violence. *Contemporary family therapy*, 38(2), 223-232.

Valla, B. (2014). *Videre, Hvordan psykiske helsetjenester kan bli bedre*. Oslo: Gyldendal akademisk forlag.
Vignoles, V. L., Owe, E., Becker, M., Smith, P. B., Easterbrook, M. J., Brown, R., . . . Cadena, M. P. (2016). Beyond the 'east–west'dichotomy: Global variation in cultural models of selfhood. *Journal of Experimental Psychology: General, 145*(8), 966.
Virkki, T., Husso, M., Notko, M., Holma, J., Laitila, A., & Mäntysaari, M. (2015). Possibilities for intervention in domestic violence: frame analysis of health care professionals' attitudes. *Journal of social service research*, 41(1), 6-24.
Vivian, D., & Malone, J. (1997). Relationship factors and depressive symptomatology associated with mild and severe husband-to-wife physical aggression. *Violence and Victims, 12*(1), 3.
Vocks, S., Tuschen- Caffier, B., Pietrowsky, R., Rustenbach, S. J., Kersting, A., & Herpertz, S. (2010). Meta- analysis of the effectiveness of psychological and pharmacological treatments for binge eating disorder. *International Journal of Eating Disorders, 43*(3), 205-217.
Von Bertalanffy, L. (1972). The history and status of general systems theory. *Academy of management journal, 15*(4), 407-426.
Von Glasersfeld, E. (1988). Constructivism as a Scientific Method. *Scientific Reasoning Research Institute Newsletter, 3*(2), 8-9.
Waddell, G., & Burton, A. K. (2006). *Is work good for your health and well-being?* : The Stationery Office.
Walfish, S., McAlister, B., O'donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports, 110*(2), 639-644.
Walker, L. E. (1979). *The battered woman*. New York: Harper and Row.
Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work*: Routledge.
Warr, P. (2003). Well-Being and the Workplace. *Well-being: Foundations of hedonic psychology*, 392.
Whisman, M. A., & Uebelacker, L. A. (2006). Impairment and distress associated with relationship discord in a national sample of married of cohabitating adults. *Journal of Family Psychology, 20(3)*, 369-371.
Willig, C. (2019). Ontological and epistemological reflexivity: A core skill for therapists. *Counselling and Psychotherapy Research*.
Wilson, G. T., Wilfley, D. E., Agras, W. S., & Bryson, S. W. (2010). Psychological treatments of binge eating disorder. *Archives of General Psychiatry, 67*(1), 94-101.
Wolf, K. A., & Foshee, V. A. (2003). Family violence, anger expression styles, and adolescent dating violence. *Journal of Family Violence, 18*(6), 309-316.
Wood, A. M., White, I. R., & Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical trials, 1*(4), 368-376.

World Health Organization. (2015). Violence. Retrieved from http://www.who.int/topics/violence/en/

World Health Organization. (2017, November 2017). Violence against women. Intimate partner and sexual violence against women. Retrieved from http://www.who.int/mediacentre/factsheets/fs239/en/

Young, M., Luquis, R., Denny, G., & Young, T. (1998). Correlates of sexual satisfaction in marriage. *The Canadian Journal of Human Sexuality, 7*(2), 115-128.

Zahl-Olsen, R., & Oanes, C. J. (2017). An Anthill of Questions that Made Me Prepare for the First Session: A Clinical Vignette of the Usage of STIC Feedback System. In T. Tilden & W. B. E. (Eds.), *Routine Outcome Monitoring in Couple and Family Therapy* (pp. 189-209). Cham, Switzerland: Springer.

Zinbarg, R. E., Pinsof, W. M., Quirk, K., Kendall, A., Goldsmith, J., Hardy, N., . . . Latta, T. (2018). Testing the convergent and discriminant validity of the Systemic Therapy Inventory of Change Initial scales. *Psychotherapy Research, 28*, 734-749. doi:https://doi.org/10.1080/10503307.2017.1325022

Øverlien, C. (2012). *Vold i hjemmet - barns strategier*. Oslo: Universitetsforlaget.

Øverlien, C., & Holt, S. (2018). Letter to the Editor: Research on Children Experiencing Domestic Violence. *Journal of Family Violence*. doi:10.1007/s10896-018-9997-9

10 Appendix, paper 1-3

Paper 1

Zahl-Olsen, R., Gausel, N., Håland, Å.T., Tilden, T. (In review). Monitoring therapeutic change through diversity or simplicity? A conservative, critical Confirmatory Factor Analysis test of a Routine Outcome Monitoring system.

Submitted to *Frontiers of Psychology* April 4 2019

Monitoring therapeutic change through diversity or simplicity? A conservative, critical Confirmatory Factor Analysis test of a Routine Outcome Monitoring system

1 Introduction

Even though psychotherapy has been found to be effective for a large portion of those people seeking help (e.g., Lambert, 2007; Sexton, Datchi, Evans, LaFollette, & Wright, 2013), about half of the clients do not benefit from therapy, based on the measures used. In fact, some even get worse (Ogles, 2013). In an attempt to better understand, and explain, why some clients do not benefit from therapy while others do, an empirical approach that dvelves into the dynamics behind the psychotherapeutic process has evolved over the last twenty years, named Routine Outcome Monitoring systems (ROM); (Lambert, 2010; Ogles, 2013; Tilden & Wampold, 2017). Because ROM's intention is to monitor the outcome of the psychotherapeutic process as it evolves over time, it has provided valuable insights and help for both the client and the therapist (Duncan, Miller, & Sparks, 2004; Olkowska, Sundet, & Karlsson, 2018; Valla, 2014; Zahl-Olsen & Oanes, 2017). Moreover, some studies suggest that the use of ROM also improves the efficiency of the treatment (Anker, Duncan, Sparks, & La Greca, 2009; Reese, Toland, Slone, & Norsworthy, 2010; Tilden & Wampold, 2017) and that they identify cases that do not develop as expected in real time (Lambert & Shimokawa, 2011). However, recent studies have found no difference in outcomes whether the ROM has been used or not (Kendrick et al., 2016; Tilden et al., 2019). One reason for this discrepancy might be that some ROM's are causing confusion rather than clarity because they oversimplify reality (Tasma et al., 2016; Zahl-Olsen & Oanes, 2017).

It is crucial for therapists and researchers to understand both the client and the ongoing therapeutic process to give them the appropriate treatment at the right time in the process. The question is how best to understand the client and the therapeutic process. There has been an ongoing discussion within the field of psychotherapy (Hubble, Duncan, & Miller, 1999; Wampold & Imel, 2015) whether this is best achieved through a focus on more general, overarching questions of wellbeing (i.e., common factors) such as: 'How are you doing, individually?' (e.g., Hubble et al., 1999), or whether it is best achieved through a focus on more specific factors tapping into wellbeing (e.g., Lambert et al., 1996; Pinsof et al., 2009; Stratton et al., 2014) with detailed questions like: 'Have you felt sad most of the day?'. Naturally, the more specific you want to be, the more specific questions you would have to ask in order to cover the different areas of interest. Thus, in a ROM focusing on specific factors, each area of interest becomes one of many subscales. Hence, the use of such a ROM has the potential to yield information from comprehensive areas of life.

Because interpersonal psychological mechanisms and relational interactions are diverse (Schibbye, 2009; Wampold & Imel, 2015), therapists tend to be concerned that the complexity of some of these ROM measures is too time-consuming for the psychotherapeutic process (Boswell, Kraus, Miller, & Lambert, 2015; Miller, Duncan, Brown, Sparks, & Claud, 2003). Because of the complexity and time burden, many therapists prefer a less diverse and therefore simplified version of the ROMs (Brown, Dreis, & Nace, 1999; Oanes, Borg, & Karlsson, 2015). The argument is often made that the shorter ROMs are easier to administer and that they are almost as good as, or as good as, the longer versions (see, e.g., O'Hanrahan et al., 2017). However, Zinbarg and colleagues (2018) argue that a measure

with several subscales will provide richer, and more helpful information for the clinician than a single- scaled general measure will do. Several of the comprehensive ROMs present their results in subscales and total scales. The subscales are theorized to capture the details while the total scale is supposed to give a more general understanding. These two levels of simplicity are the same for most of the ROMs, e.g. the Outcome Questionnaire (OQ-45; Lambert et al., 1996), the Systemic Clinical Outcome and Routine Evaluation (SCORE 15; Carr & Stratton, 2017; Stratton et al., 2014), The Treatment Outcome Package (TOP; Kraus, Seligman, & Jordan, 2005) and the Systemic Therapy Inventory of Change (STIC; Pinsof et al., 2009; Pinsof et al., 2015). Hence, it might seem like the comprehensive measures are intended to meet both the need for details that Zinbarg and colleagues (2018) argue for and the simplicity asked for by therapists. This situation is also explicitly mentioned by some of the developers as the reason for including both subscales and total scales (see. e.g., Lambert et al., 1996; Pinsof, 2017).

Naturally, assessing therapeutic change as experienced by clients, must be simplified to a certain extent when using a ROM. However, we must be willing to ask; How much simplification of the questions asked to the clients can be applied before it no longer represents the diverse reality of the psychotherapeutic process? We believe the answer to this question is: As little as possible. We will test this hypothesis on one of these ROM's, the Systemic Therapy Inventory of Change (STIC), and assess if it represents the responses of the clients better if it is allowed to remain diverse, than if the same ROM is simplified.

The STIC system is a comprehensive ROM that builds on a multi-systemic and multi-dimensional perspective of family therapy (Pinsof et al., 2009). It consists of six scales: Individual Problems and Strengths (IPS), Family of origin, Relationship with partner, Family/household, Child problems and strengths, and Relationship with child. All six scales have several subscales (from three to eight). However, even if STIC is comprehensive, and might seem overwhelming in details for those who prefer one single score, the developers of the STIC tried to present the results in one comprehensive and one simple way. The first one is the total scales model (e.g., IPS) presenting one simple graph or one single number presenting the level of distress. The other model is the sub-scales model, presenting all the sub-scales within the total scale as different graphs or numbers indicating the level of distress for each of the subscales. For the IPS scale, the eight sub-scales load from two to six items each. Since the STIC system provides both a diverse and a simplified model, we wanted to test what gives the best picture of the diverse reality of a client. Hence, we wanted to compare, and thereby assess, the two different models presented by the STIC system: the total scale model, and the subscales model. We chose to do so by testing the IPS scale on a Norwegian sample of clients.

2 The Current Study

The data in this study stems from Norway where couple and family therapy treatment is provided to the public in a stepped level of care. The first level of care, where no referral is needed, is represented by two agencies. The second level of care is represented by an outpatient agency, where a referral is needed. The third and final level of care is represented by an inpatient agency where referral is needed. Since the data ordinates from ordinary clinical practice, no inclusion or exclusion criteria have been used, except for the criteria each site has for accepting clients for treatment. The data stems from clients of more than 40 therapists at all three levels of the Norwegian couple and family therapy treatment.

2.1 Participants and Procedure

The sample consists of 841 clients (51.8% women; Mean age: 40; Age range: 12-72; see Table 1 for further descriptives) recruited from four different couple and family therapy sites. Data were collected from March 2010 to April 2016 – prior to or at the start of therapy – and they all completed the STIC start questionnaires during a pilot (Tilden, Håland, Hunnes, Fossli, & Oanes, 2015) and an RCT study (Tilden et al., 2019).

2.2 Ethics

Written informed consent for collecting the project data was obtained from each participant. For participants under the age of 16, written informed consent was obtained from the parents or guardians of participants. This study was approved by the Modum Bad Ombudsman for Data Protection and the Regional Ethics Committee for Medical Research with human subjects (2017/96). The primary study is registered at ClinicalTrials.gov.

insert table 1, Descriptive statistics about here

2.3 Measures

The STIC has six scales, and we have chosen to use the Individual Problems and Strength (IPS) scale of the STIC system. We chose to analyze only one scale, firstly, because most other ROMs only have one total scale and secondly, since that scale is directly assessing the respondents' experience of themselves. The IPS scale consists of eight latent subscales loading a total of 22 items as described in table 2. The subscales and their corresponding Cronbach's alphas are in our sample, at start of treatment: Flexibility/resilience ($\alpha = .67$), Life functioning ($\alpha = .78$), Open expression ($\alpha = .68$), Self-acceptance ($\alpha = .71$), Disinhibition ($\alpha = .53$), Negative affect ($\alpha = .86$), Self-misunderstanding ($\alpha = .66$), and Substance abuse ($\alpha = .20$). The IPS total scale ($\alpha = .89$) as measured as a single, omnibus latent variable containing 22 items as described by the creator of the system (Pinsof et al., 2009).

insert table 2, All items and subscales in the IPS scale about here

2.4 Statistical approach

We used AMOS 25 in order to test our hypothesis in a Confirmatory Factor Analysis (CFA) with maximum likelihood estimation. We adapted the conservative approach to scale validation as suggested by Gausel, Leach, Vignoles, and Brown (2012) and Gausel, Vignoles and Leach (2016); first testing the hypothesized preferability of the diverse model, then contrasting this result against a collapsed, omnibus model in order to test their fit up against each other. In line with the approach suggested by Gausel et al., (2012; 2016), no items were allowed to cross-load on any of the factors, and no error terms were allowed to correlate. Naturally, the latent factors were allowed to correlate.

Given the complexity of the models, we calculated other fit-statistics than just the chi-square (Kline, 2016); we calculated the $\chi 2$/df ratio and several chi-square-based fit indices (incremental fit index [IFI], comparative fit index [CFI]). In, addition, we assessed the good model fit was by our observation of a residual index (root-mean-square error of approximation [RMSEA]; see Kline, 2016). RMSEA is also presented with 90% confidence intervals as suggested by Kline (2016). To assess how well the total-scale model fit the data compared to the models with subscales we

calculated the Δ $\chi 2$ and Δ AIC statistics. The AIC statistics are based on chi-square calculations but take model complexity into account. Models with a lower AIC are superior, because they better represent the covariance between the variables examined (see Kline, 2016) such that when investigating AIC, it is not the absolute size of the statistics that matter, but the relative difference between the values where the lower value is the superior one (Kline, 2016).

3 Results

The diverse model. The diverse measurement model, originally proposed by Pinsof and colleagues (2009; 2015), fit the data well, especially when considering the complexity of the model, $\chi 2$ (181) = 605,994, p <.001, $\chi 2$/df ratio = 3.35, IFI = .936, CFI = .935, AIC = 793.994, RMSEA = .053 [.048 - .058]. As seen in Figure 1, all factor loadings were statistically significant (all p's < .001) with factor loadings ranging from standardized λ = .30 to .86, where most standardized λ were .55 or higher. These outcomes indicate that all factors were well defined by their respective items (see Gausel et al., 2012; 2016). Moreover, Figure 1 also illustrates the variation in the correlations among the eight different factors (ranging from r = .06, p = .398, to r = .89, p < .001), confirming that the diverse model presents eight different, but correlated constructs (Gausel et al., 2012; 2016) each representing different, meaningful aspects of the client's life (Pinsof et al., 2009). We checked all the factors with a correlation of .70 and above for multicollinearity. As a rule of thumb, a VIF >10 indicates problematic multicollinearity (Field, 2018). The VIF was not above 1.8 among any of the factors in our model and, thus, well below the suggested cut-off values for multicollinearity.

insert figure 1 about here

The simplified, omnibus model. The second model presented in the STIC system (Pinsof et al., 2009; 2015) is a measurement model where all the twenty-two items load onto a single factor (i.e., the IPS total). We, therefore, tested this one-factor model with a second step, contrasting its fit up against the diverse model as suggested by Gausel and colleagues (2012; 2016). As expected, its fit was poor, $\chi 2$ (209) = 1949,931, p <.001, $\chi 2$/df ratio = 9.33, IFI = .736, CFI = .734, AIC = 2081,931, RMSEA = .100 [.096 - .104]. Naturally, having only one factor, there are no correlations to investigate, but as seen in Figure 2, factor loadings ranged from as little as standardized λ = .17 to .78 (all p's < .001), where only 11 of the 22 items had a factor loading larger than standardized λ .55 which is considered a minimum level for a good factor loading (Comrey & Lee, 2016; Tabachnick & Fidell, 2014).

insert figure 2 about here

Importantly, when comparing the chi-square value of the two models, our diverse model fit data significantly better than the simplified, omnibus model; Δ χ^2 (28) = 1343.937, p < .001. Moreover, the difference in AIC (Δ AIC = 1287.937) between the two models assured us that the diverse model fit data much better than the simplified, omnibus model. Hence, the diverse model proved to be superior to the simplified, omnibus model.

Other possible models. Looking at the four correlations in Figure 1 that are higher or equal to r = .70, some might wonder if the diverse model would fare better if it were just a little less complex. Hence, we tested three new models, each collapsing the factors correlating equal to or above r = .70 and compared it to the diverse model. We also tested a fourth model collapsing all three solutions at

once and then comparing this result to the diverse model. As expected, the diverse model proved to be superior to all these four alternative models. First, it fits better than a model collapsing "open expression" with "self-acceptance", $\Delta \chi^2$ (7) = 149.44, $p < .001$ and Δ AIC = 135.440. Second, it fits better than a model collapsing "Flexibility/resilience" with "self-acceptance", $\Delta \chi^2$ (7) = 27.761, $p < .001$ and Δ AIC = 13.761. Third, it fits better than a model collapsing "Flexibility/resilience" with "self-misunderstanding", $\Delta \chi^2$ (7) = 75.735, $p < .001$ and Δ AIC = 61.735. Forth, it fits better than a model collapsing "self-misunderstanding" with "self-acceptance", $\Delta \chi^2$ (7) = 95.713, $p < .001$ and Δ AIC = 81.713. Fifth, it fits better than a model where all of the four mentioned factors were collapsed, $\Delta \chi^2$ (18) = 301.710, $p < .001$ and Δ AIC = 260.71.

4 Discussion

The main question in this study is: How much simplification of the models can be applied before it no longer represents the diverse reality of the clients and the psychotherapeutic process? The main finding was: Nothing. As expected, the diverse model proved to be superior to a simplified, omnibus model, as well as other possible solutions.

4.1 Model fit statistics favor the diverse model

The diverse model achieved a superior fit to all other alternatives. This result means that the model supports our hypothesis that a diverse and detailed model represents the client's answers well (Zinbarg et al., 2018). All the other models performed statistically significantly worse than this diverse model. When comparing the diverse model with the simplified, omnibus model, we found the results to be not just a slightly less good fit, but seriously worse. This fact implies that the diverse model approximated the observed values much better than the simplified model. In clinical terms, this outcome means that the diverse model was better at predicting the actual answers given by the clients, while the simplified model(s) was worse at it. Indeed, the simplified model produced higher values for residuals as measured with RMSEA. In line with this result, Tabachnick and Fidell (2014) suggest that for a model to be considered a good model, it cannot produce residuals above a certain level, and the simplified, omnibus model failed on these premises. In clinical terms, this result means that the model risks becoming so simplified that it hinders the therapist from understanding the client and/or the therapeutic process. In fact, it produces so much error that it will be hard for the therapist to assess what is correct and what is not. This finding that the simplified model performed residuals (i.e., RMSEA) above the suggested threshold, is supported by the study of Zinbarg and colleagues (2018), who tested the same simplified model (i.e., IPS total scale) on a sample of 476 US clients. These authors got RMSEA values even higher than ours, indicating that it is not a good model. Consequently, our conclusion, based on the model fit statistics, is not to use the IPS total scale.

4.2 Factor loadings favoring the diverse model

When investigating the factor loadings in the simplified model, we found 11 items, or half of the list, below the suggested minimum level (Comrey & Lee, 2016; Tabachnick & Fidell, 2014). Hence, the factor loadings do not support the simplified model being a good model. However, Zinbarg and colleagues (2018) and Pinsof and colleagues (2008; 2015) argue that this simplified model of applying the STIC data is a good model that represents the clients well, and they suggest using it. On the contrary, our study found that the simplified model will not inform the therapist of important changes, even if the STIC explicitly has asked the client about several issues. E.g., if the client has had fits of rage, they could not control.

Half of the factor loadings in the simplified model do not support the simplified model, but not all factor loadings in the diverse model support the diverse model. Hence, one might wonder if the diverse model is not a good model either. Therefore, we investigated in more detail the items in the diverse model with low factor loadings. We found that from a clinical perspective, we will argue they belong in the model. One example is the item asking if the clients have had thoughts about seriously harming or killing someone. This item has a low factor loading, and, in fact, almost every client (94.1%) in our sample answered never/not at all to that question. But, since a ROM, first and foremost, is meant to be a clinically helpful tool (Pinsof & Wynne, 2000; Tilden & Wampold, 2017) we support the founder of the scale (Pinsof et al., 2009) that this item is clinically important and, thereby, should be included. We aim to discuss how much simplification can be done to a model and still have it be a good aid for psychotherapists. So far in our discussion, the simplified model has failed to perform well enough when considering the model fit statistics and the factor loadings. However, we also performed three additional steps in our assessment.

4.3 Explained variance favors the diverse model

We found that the diverse model had a reasonable to the good explained variance for all items, except for the ones just mentioned, with low factor loadings. This finding supports our hypothesis that a diverse model, as this eight t-factor model, is a good model to represent the client's answers. This finding was in line with the conclusion of Pinsof and colleagues (2015) when they analyzed the eight-factor structure of STIC on a representative US sample. However, Kline (2016) suggests that an ideal result of a CFA explains the majority (>50%) of the variance in every item. The diverse model explains ten of the items above this level while the simplified model explains only three of the 22 items above the suggested 50% level. Clinically, this fact means that the simplified model does not help the psychotherapist understand the change going on in therapy. If most of the change is explained by factors outside of the model, what use is there of the model then? We argue that when the model only explains very little of the change in many of the areas we believe is of importance to the therapeutic process, there is not much use for the model, and it might even be contraindicating. Since therapists are overly optimistic about the improvement of their clients (Walfish, McAlister, O'donnell, & Lambert, 2012) and not many clients inform their therapists about the lack of improvement (Hannan et al., 2005; Harmon et al., 2007) the last thing a therapist needs is a ROM that does not help him or her to assess the treatment.

4.4 The diverse model: Theoretical support

Both models analyzed in this study, the diverse and the simplified model, are constructed on theory (Pinsof & Wynne, 2000; Pinsof et al., 2009) and are explained to give both practical and theoretical sense (Pinsof, 2017; Pinsof et al., 2009). Thus, we want to discuss the models from a theoretical perspective. The diverse model has eight distinctly different subscales. The names of these subscales indicate that they measure different aspects of the client's life. It is logical that e.g. negative affect and life functioning are two different phenomena. The same with substance abuse and open expression. Furthermore, this logical structure was supported by the clients. E.g., the covariance between substance abuse and open expression was very low. Another example is the covariance between the subscales disinhibition and life functioning that was found to be low. This situation means that the clients are strongly saying that these subscales are not measuring the same phenomenon. What the simplified, omnibus model does is to combine them all on the same scale. As demonstrated from our data, this choice has no empirical support.

When all of the eight factors from the diverse model are combined to one single factor in the simplified, omnibus model what we do mathematically is forcing the covariance to be 1 even if our

investigation of the diverse model clearly told us that many of them are not even close to 1. That is violating the best possible fit for the data and we, therefore, suggest that the simplified, omnibus model is conceptually wrong, at least when based on these 22 items.

These theoretical and mathematical conceptual flaws apply to other ROMs, which measure diverse aspects of client's lives. It also applies to the measures that are provided in simplified versions such as SCORE 40, 28, and 15 (O'Hanrahan et al., 2017). In those cases, the developers argue that even if several items are excluded, the measure still captures the same picture. Our analysis, on STIC IPS scale, has shown that that situation is not the case. Hence, we argue that even though it might seem compelling, we would like to warn against using simplified measurement-scales as it is possible that simplification of this sort increases the measurements impreciseness, such that one risks misguiding both the clinicians and the researchers in their attempt to better understand the therapeutic change.

What all this means is that even if many ROMs are providing both diverse and simplified models (e.g., (Busby, Christensen, Crane, & Larson, 1995; Carr & Stratton, 2017; Kraus et al., 2005; Lambert et al., 1996; Pinsof et al., 2009; Pinsof et al., 2015; Stratton et al., 2014) and some shorter versions (e.g., (Carr & Stratton, 2017) we question the usefulness of the simplified models. We agree with Zinbarg and colleagues (2018) that a measure with several subscales will provide richer, and more helpful information for the clinician than any single scaled general measure will do. However, some of the discussion about ROMs (for an overview, see (Ogles, 2013; Wampold & Imel, 2015) is not about if one diverse measure can be simplified, like what we have tested in this study, but if tracking general items can be helpful as a ROM. Several studies support that general items are improving the outcomes (see, e.g., (Anker et al., 2009; Hubble et al., 1999). We have not tested any general measure in this study, but based on our results, we question the usefulness of overly simplified models, like one scale models.

5 Possible limitations

The strength of this study is the number of participants and that it includes participants from 12 to 72 years of age, from different parts of the country, clients with no prior experience with therapy and clients with several years of prior therapy and with varying levels of distress (Tilden et al., 2019). It is also a strength of the generalizability of the findings that we have included data from clients both at low threshold outpatient agencies, a medium threshold outpatient clinic, and a high threshold inpatient clinic.

The study presented in this article has been performed in the Norwegian language, and some may argue that the sample is not representative of an international population. However, Norwegians seem to be similar to other nationalities in the Western world (Becker et al., 2018). For example, the divorce rate in Norway is similar to citizens from other European countries (Eurostat, 2015). Further, it is likely that individual and relational problems are common internationally, and if differences occur, they are not as expected (Owe et al., 2013; Vignoles et al., 2016).

Finally, we argue that since the results from our analysis of the diverse model are equivalent to the ones on the representative US sample (Pinsof et al., 2015) that consisted of a diverse population with people with different ethnicities the results from this study are also applicable for international populations.

6 Conclusion and implications

One result of this study is that we have confirmed the factor structure of the STIC measure in a Norwegian sample. More importantly, we have contributed to the ongoing debate in the field of psychotherapy about whether to use simple or diverse ROM systems (see, e.g., Hubble et al., 1999; Wampold & Imel, 2015; Zinbarg et al., 2018). This study is adding to this debate because it has been investigating one diverse and one simplified model. The diverse model had a good model fit, varied covariances, acceptable to good factor loadings, and a theoretical rationale that makes sense.

In contrast, these circumstances were not found to be the case with the simplified, omnibus model. Indeed, we found no logical, conceptual, or analytical support for the omnibus model based on the same 22 items. Even if it is tempting to simplify, based on our findings, we are convinced that this simplification should not be done.

Possible implications for clinicians using STIC in their clinical practice, trying to understand clients in the therapy session, are, therefore, that they should be cautious in the use of total scales in their clinical evaluation of therapy developments since it might have them misunderstand their clients, rather than to understand them. Thus, the following treatment interventions may be misguided. A possible broader implication is that caution might be taken for the simplification of other ROMs, which measure diverse aspects of client's lives as well.

For researchers investigating groups of clients, it is important to consider that the IPS total scale is a measure that averages a diverse set of concepts in clients' lives. Hence, analysis of change using the total score could imply that the clients experienced a change in very different aspects. Hence, we suggest an assessment of the severity of problems or change in therapy using the sub-scales rather than the total score. In general, researchers should be cautious in the use of a shorter version of a comprehensive measure, or by using a total scale in their analysis, instead of sub-scales.

Based on our findings, we do not support using all the items in IPS to construct one single measure for individual problems and strengths. It might be possible to construct a meaningful and reliable IPS total scale from some of these questions, but that is outside the scope of this study. Maybe it is even better to construct new questions to consider that factor.

Our bottom line conclusion is that diverse therapeutic experiences should be allowed to be diverse.

7 References:

Anker, M. G., Duncan, B. L., Sparks, J. A., & La Greca, A. M. (2009). Using Client Feedback to Improve Couple Therapy Outcomes: A Randomized Clinical Trial in a Naturalistic Setting. Journal of Consulting and Clinical Psychology, 77(4), 693-704. doi:10.1037/a0016062

Becker, M., Vignoles, V. L., Owe, E., Easterbrook, M. J., Brown, R., Smith, P. B., . . . Torres, A. (2018). Being oneself through time: Bases of self-continuity across 55 cultures. Self and Identity, 17(3), 276-293.

Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. Psychother Res, 25(1), 6-19. doi:10.1080/10503307.2013.817696

Brown, J., Dreis, S., & Nace, D. K. (1999). What really makes a difference in psychotherapy outcome? Why does managed care want to know? The Heart and Soul of Change: What Works in Therapy (pp. 389 - 406).

Busby, D. M., Christensen, C., Crane, D. R., & Larson, J. H. (1995). A revision of the dyadic adjustment scale for use with distressed and nondistressed couples: construct hierarchy and multidimensional scales. Journal of Marital and Family Therapy, 21(3), 289-308. doi:10.1111/j.1752-0606.1995.tb00163.x

Carr, A., & Stratton, P. (2017). The SCORE family assessment questionnaire: a decade of progress. Family Process, 56(2), 285-301.

Comrey, A. L., & Lee, H. B. (2016). A first course in factor analysis (2nd. ed.). NY: Routledge.

Duncan, B. L., Miller, S. D., & Sparks, J. (2004). The heroic client : a revolutionary way to improve effectiveness through client-directed, outcome-informed therapy (Rev. ed. ed.). San Francisco: Jossey-Bass.

Eurostat. (2015) Marriage and divorce statistics.

Field, A. (2018). Discovering statistics using SPSS, 5th edition (5th ed.). London: Sage publications.

Gausel, N., Leach, C. W., Vignoles, V. L., & Brown, R. (2012). Defend or repair? Explaining responses to in-group moral failure by disentangling feelings of shame, rejection, and inferiority. Journal of personality and social psychology, 102(5), 941.

Gausel, N., Vignoles, V. L., & Leach, C. W. (2016). Resolving the paradox of shame: Differentiating among specific appraisal-feeling combinations explains pro-social and self-defensive motivation. Motivation and Emotion, 40(1), 118-139.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. Journal of Clinical Psychology, 61(2), 155-163. doi:10.1002/jclp.20108

Harmon, S. C., Lambert, M. J., Smart, D. W., Hawkins, E. J., Nielsen, S. L., & Slade, K. (2007). Enhancing outcome for potential treatment failures: Therapist/patient feedback and clinical support tools. Psychotherapy Research, 17, 380-391.

Hubble, M. A., Duncan, B. L., & Miller, S. D. (1999). The heart and soul of change: What works in therapy: American Psychological Association.

Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., . . . de Jong, K. (2016). Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. Cochrane Database of Systematic Reviews(7).

Kline, R. B. (2016). Principles and practice of structural equation modeling (4th ed.). New York: Guilford publications.

Kraus, D. R., Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. Journal of Clinical Psychology, 61(3), 285-314.

Lambert, M. J. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. . Psychotherapy Research, 17, 1-14.

Lambert, M. J. (2010). Yes, it is time for clinicians to routinely monitor treatment outcome.

Lambert, M. J., Burlingame, G., Umphress, V., Hansen, N., Vermeersch, D., & Clouse, G. (1996). The reliability and validity of the Outcome Questionnaire. Clin Psychol Psychother, 3, 249 - 258.

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. Psychotherapy, 48, 72 - 79.

Miller, S., Duncan, B. L., Brown, J., Sparks, J., & Claud, D. (2003). The outcome rating scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. J Brief Ther, 2, 91 - 100.

O'Hanrahan, K., Daly White, M., Carr, A., Cahill, P., Keenleyside, M., Fitzhenry, M., . . . O'Shea, H. (2017). Validation of 28 and 15 item versions of the SCORE family assessment questionnaire with adult mental health service users. Journal of Family Therapy, 39(1), 4-20.

Oanes, C. J., Borg, M., & Karlsson, B. (2015). Significant Conversations or Reduced Relational Capacity? Exploring Couple and Family Therapists' Expectations for Including a Client Feedback Procedure. Australian and New Zealand Journal of Family Therapy, 36(3), 342-355. doi:10.1002/anzf.1110

Ogles, B. M. (2013). Measuring change in psychotherapy research. In M. J. Lambert (Ed.), Bergin and Garfield's Handbook of psychotherapy research and behavior change 6th ed. (pp. 134-166). New Jersey: Wiley.

Olkowska, A., Sundet, R., & Karlsson, B. E. (2018). Kan klient- og resultatstyrt praksis (KOR) bidra til økt brukermedvirkning og med dette til bedrings- og mestringsprosesser i terapi? Fokus på familien, 46(04), 294-315. doi:10.18261/issn.0807-7487-2018-04-05

Owe, E., Vignoles, V. L., Becker, M., Brown, R., Smith, P. B., Lee, S. W., . . . Gheorghiu, M. (2013). Contextualism as an important facet of individualism-collectivism: Personhood beliefs across 37 national groups. Journal of Cross-Cultural Psychology, 44(1), 24-45.

Pinsof, W. M. (2017). The Systemic Therapy Inventory of Change—STIC: A Multi-systemic and Multi-dimensional System to Integrate Science into Psychotherapeutic Practice Routine Outcome Monitoring in Couple and Family Therapy (pp. 85-101): Springer.

Pinsof, W. M., & Wynne, L. C. (2000). Toward progress research: Closing the gap between family therapy practice and research. Journal of Marital and Family Therapy, 26(1), 1-8.

Pinsof, W. M., Zinbarg, R. E., & Knobloch-fedders, L. M. (2008). Factorial and Construct Validity of the Revised Short Form Integrative Psychotherapy Alliance Scales for Family, Couple, and Individual Therapy. Family Process, 47(3), 281-301. doi:10.1111/j.1545-5300.2008.00254.x

Pinsof, W. M., Zinbarg, R. E., Lebow, J., Knobloch-Fedders, L. M., Durbin, E., Chambers, A., . . . Friedman, G. (2009). Laying the foundation for progress research in family, couple, and individual therapy: The development and psychometric features of the initial Systemic Therapy Inventory of Change. Psychotherapy Research, 19(2), 143-156.

Pinsof, W. M., Zinbarg, R. E., Shimokawa, K., Latta, T. A., Goldsmith, J. Z., Knobloch-Fedders, L. M., . . . Lebow, J. L. (2015). Confirming, Validating, and Norming the Factor Structure of Systemic Therapy Inventory of Change Initial and Intersession. Family Process, 54(3), 464-484. doi:10.1111/famp.12159

Reese, R., Toland, M., Slone, N., & Norsworthy, L. (2010). Effect of client feedback on couple psychotherapy outcomes. Psychother Theory Res Pract Train, 47, 616 - 630.

Schibbye, A.-L. L. (2009). Relasjoner: et dialektisk perspektiv på eksistensiell og psykodynamisk psykoterapi: Universitetsforlaget.

Sexton, T. L., Datchi, C., Evans, L., LaFollette, J., & Wright, L. (2013). The effectiveness of couple and family-based clinical interventions. In M. J. Lambert (Ed.), Bergin and Garfield's Handbook of psychotherapy and behavior change, 6th edition. (6th edition ed., pp. 587-639). New Jersey: John Wiley & Sons.

Stratton, P., Lask, J., Bland, J., Nowotny, E., Evans, C., Singh, R., . . . Peppiatt, A. (2014). Detecting therapeutic improvement early in therapy: validation of the SCORE-15 index of family functioning and change. Journal of Family Therapy, 36(1), 3-19.

Tabachnick, B. G., & Fidell, L. S. (2014). Using Multivariate Statistics, 6th edt. Harlow: Pearson Education Limited.

Tasma, M., Swart, M., Wolters, G., Liemburg, E., Bruggeman, R., Knegtering, H., & Castelein, S. (2016). Do routine outcome monitoring results translate to clinical practice? A cross-sectional study in patients with a psychotic disorder. BMC Psychiatry, 16(1), 107.

Tilden, T., Håland, Å. T., Hunnes, K., Fossli, G., & Oanes, C. J. (2015). Utprøving av systematisk tilbakemelding i par- og familieterapi: Barrierer og utfordringer. . Fokus på familien, 43, 292-312.

Tilden, T., Wampold, B., Ulvenes, P. G., Zahl-Olsen, R., Hoffart, A., Barstad, B., . . . Håland, Å. T. (2019). Feedback in Couple and Family Therapy: A Randomized Clinical Trial. Family Process.

Tilden, T., & Wampold, B. E. (2017). Routine Outcome Monitoring in Couple and Family Therapy: The Empirically Informed Therapist: Springer.

Valla, B. (2014). Videre, Hvordan psykiske helsetjenester kan bli bedre. Oslo: Gyldendal akademisk forlag.

Vignoles, V. L., Owe, E., Becker, M., Smith, P. B., Easterbrook, M. J., Brown, R., . . . Cadena, M. P. (2016). Beyond the 'east–west'dichotomy: Global variation in cultural models of selfhood. Journal of Experimental Psychology: General, 145(8), 966.

Walfish, S., McAlister, B., O'donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. Psychological Reports, 110(2), 639-644.

Wampold, B. E., & Imel, Z. E. (2015). The great psychotherapy debate: The evidence for what makes psychotherapy work: Routledge.

Zahl-Olsen, R., & Oanes, C. J. (2017). An Anthill of Questions that Made Me Prepare for the First Session: A Clinical Vignette of the Usage of STIC Feedback System. In T. Tilden & B. Wampold (Eds.), Routine Outcome Monitoring in Couple and Family Therapy (pp. 189-209). Cham, Switzerland: Springer.

Zinbarg, R. E., Pinsof, W. M., Quirk, K., Kendall, A., Goldsmith, J., Hardy, N., . . . Latta, T. (2018). Testing the convergent and discriminant validity of the Systemic Therapy Inventory of Change Initial scales. Psychotherapy Research, 28, 734-749. doi:https://doi.org/10.1080/10503307.2017.1325022

Table 1

Descriptive statistics

	N= 841
Variable	% (N) / M [SD]
Sex: Female	51.8 (436)
Age	40.0 [8.89]
Education	
Low	51.8 (436)
Medium (Bachelor)	31.2 (262)
High (Master og PhD)	15.0 (126)
Relationship status	
Comitted relationship - not married	19.3 (162)
Married	70.2 (590)
Medication	
Using medication (Some more than one medication)	32.5 (273)
Depression	13.6 (114)
Anxiety	6.9 (58)
Concentration difficulties/hyperactivity	2.6 (22)
Bipolar	3.3 (28)
Other	6.1 (51)
Prior experience with therapy	
None	25.2 (215)
Less than one year	37.2 (313)
One to three years	20.2 (170)
More than three years	16.3 (137)

Table 2

All items and subscales in the IPS scale

Subscale	Item	Item no.
FLEXIBILITY/RESILIENCE		
	How easy is it for you generally to overcome difficulties?	IPS 01
	When what I'm trying doesn't work out, I can change my approach or my plans	IPS 25
	When I get upset, I find healthy ways to make myself feel better	IPS 26
LIFE FUNCTIONING		
	Performing work/school/household tasks	IPS 02
	Managing day-to-day life	IPS 03
OPEN EXPRESSION		
	I can openly express my feelings.	IPS 21
	I can speak up for myself when the situation calls for it	IPS 22
SELF-ACCEPTANCE		
	I can be myself in every situation	IPS 23
	I am comfortable with who I am	IPS 24
DISINHIBITION		
	Thought about seriously harming or killing someone	IPS 11
	Had fits of rage you could not control	IPS 12
	Had urges or impulses that you could not control	IPS 13
NEGATIVE AFFECT		
	Had thoughts or images over and over again that you could not get rid of	IPS 05
	Felt tense or anxious	IPS 06
	Felt sad most of the day	IPS 07
	Thought about ending your life	IPS 08
	Felt hopeless about the future	IPS 09
	Not enjoyed things as much as you used to	IPS 10
SELF-MISUNDERSTANDING		
	I don't understand why I do the things I do	IPS 27
	It's tough for me to know what I'm feeling	IPS 28
SUBSTANCE ABUSE		
	Drank too much alcohol	IPS 14
	Used illegal drugs/misused prescribed medication	IPS 15

Figure 1

The diverse model

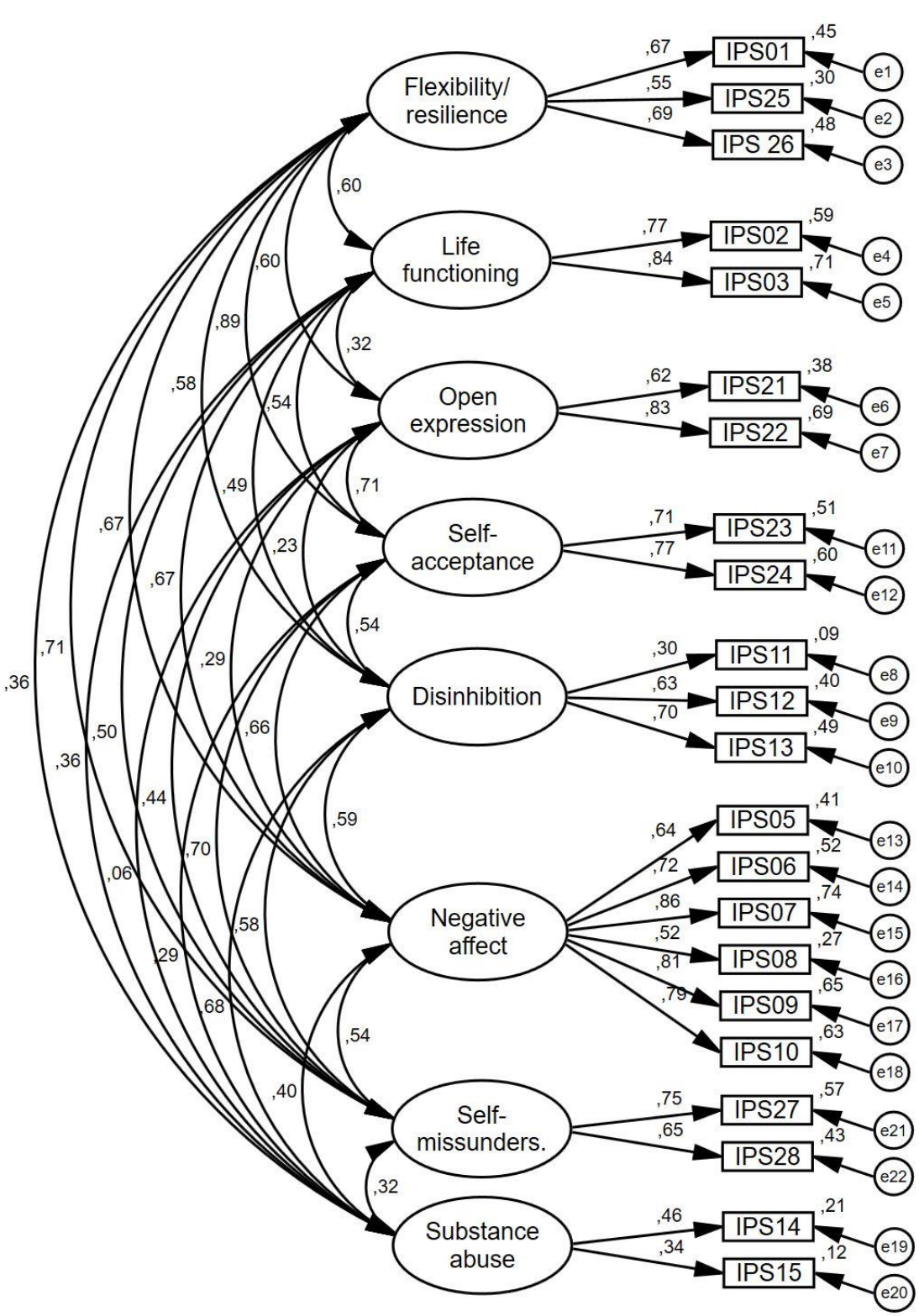
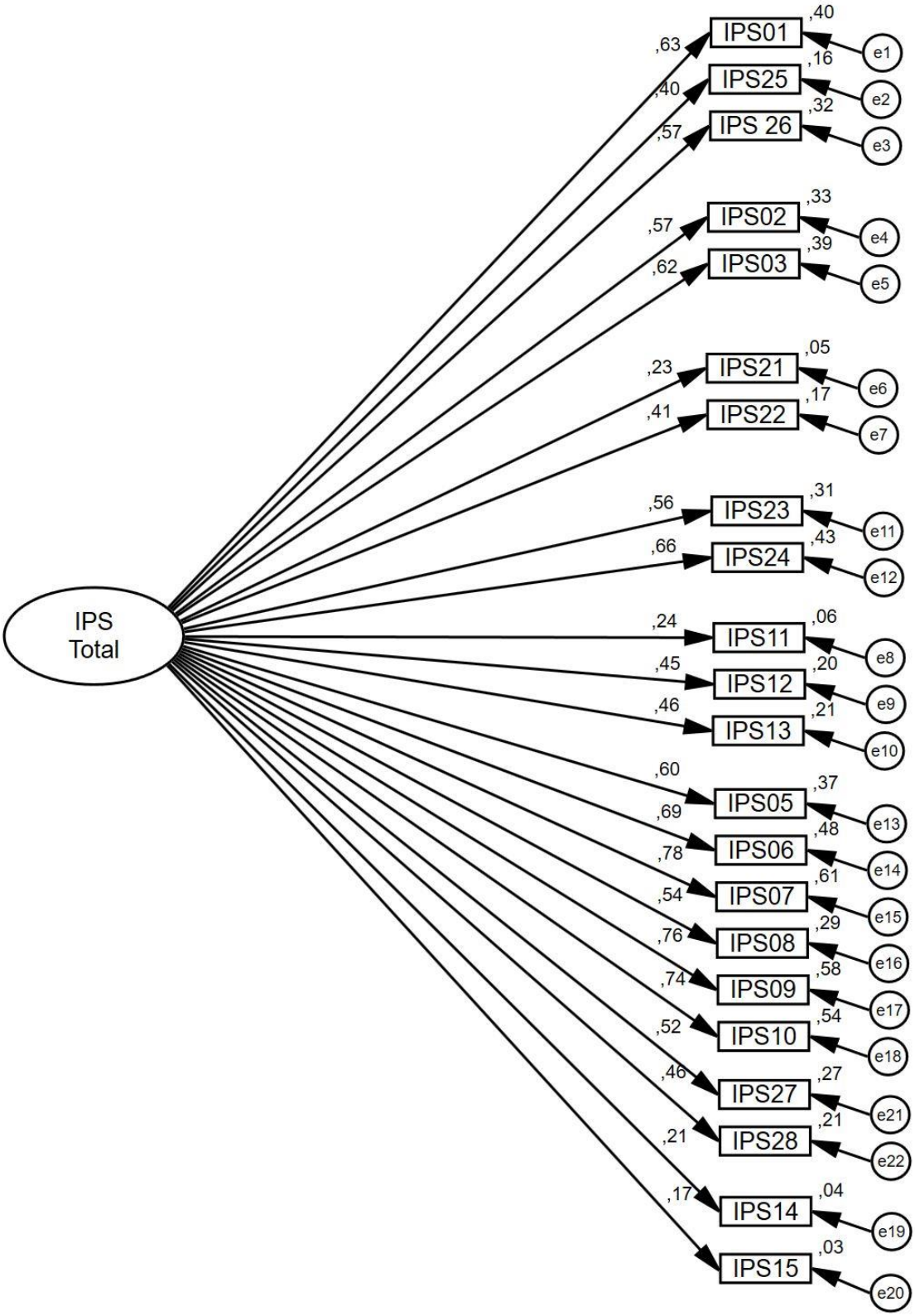


Figure 2

The simplified, omnibus model



Paper 2

Zahl-Olsen, R., Gausel, N., Zahl-Olsen A., Bjerregaard Bertelsen T., Håland Å.T., Tilden T. (2019). Physical couple and family violence among clients seeking therapy: identifiers and predictors. *Frontiers in Psychology, section Psychology for Clinical Settings. doi: 10.3389/fpsyg.2019.02847*

frontiers

Physical couple and family violence among clients seeking therapy: identifiers and predictors

Introduction

Partner violence is one of the most hurtful and traumatic experiences a human can experience as it shatters interpersonal trust and sense of safety (Stacey, Hazlewood & Shupe, 1994). Disturbingly, the prevalence of partner violence in the population has proven to be very large. An international population survey of physical abuse including 48 countries found that as many as 69 % of women state that they have been subjected to physical abuse by a partner at some time of their life, but the results varied widely by country (Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002). Although violent episodes are often thought of as a clear-cut case with a powerful perpetrator and a powerless victim, most violence is to some extent reciprocal (Gausel, Leach, Mazziotta & Feuchte, 2018) and in fact, several studies have found that as many males as females are exposed to partner violence (Capaldi et al., 2012; Haaland, Clausen, & Schei, 2005; Jose & O'Leary, 2009; Thoresen & Hjemdal, 2014). Indeed, studies from the US have found that up to 61% of clients seeking couple and family therapy (CFT) have experienced couple violence (Cascardi, Langhinrichsen, & Vivian, 1992; Jouriles & O'leary, 1985; Vivian & Malone, 1997) and it is reasonable to believe that these numbers are relevant for other comparable countries as well, such as Norway (Ormhaug, Jensen, Hukkelberg, & Egeland, 2012). However, current studies have solely focused on couple violence and did not investigate the family violence that most often co-occurs and includes children. Hence, in the present study, we wanted to investigate both the physical couple violence and the physical family violence within a CFT sample.

In intimate relationships, such as family relations and couple relationships, there are extensive opportunities of experiencing and committing moral failures. Some of them are minor (e.g., forgetting important appointments), but others are more severe and some even illegal (e.g., abuse and violence). Naturally, committing acts of violence within families and couple relationships represents a grave moral failure (Gausel et al., 2018) that seriously question your integrity as a moral person (Gausel & Leach, 2011). However, having been violent might lead the perpetrator to be concerned for their social-image as a moral person, and thus fear being condemned by others, especially if the use of violence is in risk being exposed to these others (Gausel & Leach, 2011). According to Gausel and colleagues (2011; 2018) this should promote even more violence in order to hinder the victim from exposing it or telling others about it. However, being victim to violence is a traumatic experience as well (Stacey et al., 1994) as it seriously question who you are as a person, especially your worth as a fellow human (Loring, 1994). Due to this, it is often appraised as a result of deep disrespect and because of this, it often promotes reciprocal violence (Gausel et al., 2018). In support of this, Gausel and colleagues (2018) conducted a field experiment in Liberia, Africa, with survivors of civil-wars that had been associated with groups that had been both victims and perpetrators of grave violence. In this study they argued for, and found, that when these survivors were encouraged to reflect on their victim-episodes they were significantly more motivated in seeking revenge than if they were encouraged to reflect on their perpetrator episodes. Hence, we find this especially useful in our context of couple and family violence (ie., domestic violence) as the involved parties are both perpetrators and victims of violence, a type of violence that is suggested to be the most prevalent type of domestic violence (Fusco & Fantuzzo, 2009; Stith, McCollum, & Rosen, 2011).

Even though couple violence primarily occurs among adults, it affects and includes the children severely. In fact, a Norwegian national survey (Haaland et al., 2005) found that 30% of the children witnessed the couple violence that had occurred, while an American study (Fusco & Fantuzzo, 2009) found that as many as 95% of the children had been exposed to violence exercised within the family. Of these children, 75% had an active role in trying to influence the situation by contacting a neighbor, calling the police or protecting the victims of violence with their own bodies. Being a witness to violence profoundly affect children (Apple & Holden, 1998; Kimball, 2016; Slep & O'leary, 2005; Øverlien & Holt, 2018). In line with this, a meta-analysis by Kitzmann and colleagues (2003) concludes that 63% of children exposed to violence develop internalizing (e.g., posttraumatic stress disorder) and externalizing (e.g., aggression) problems. Hence, this group of children could be expected to become clients in children and adolescence clinics. Furthermore, children who experience violence in their childhood have higher risks of using violence themselves (Raundalen, 2009) and being exposed to violence in their adult lives (Renner & Slack, 2006; Wolf & Foshee, 2003; Øverlien, 2012). Hence, there is a high risk that the problem of violence gets passed from one generation to the next.

Obviously, therapists cannot offer appropriate treatment for couple violence or family violence if it is not detected, and unfortunately research shows that clients often do not inform their therapist about violence when they seek help (Ehrensaft & Vivian, 1996; Middelborg & Samoilow, 2014). Hence, several authors have recommended the use of universal screening for couple violence. Nevertheless, most therapists do not adhere to these recommendations in their clinical practice (Schacht, Dimidjian, George, & Berns, 2009; Todahl & Walters, 2011). We have not found anyone suggesting universal screening for family violence although it occurs so frequently. The lack of such screening implies that violence is often not discovered and thereby not treated (Stith et al., 2011). This seems to be the case also when a child is the index patient. Reigstad, Jørgensen, & Wichstrøm (2006) found that 60% of the children in their study had experienced physical violence in their family, but that was reported in the journaling system only in 0.4% of these cases. Thus, it was not discovered or not found to be important for the treatment. The lack of couple and family violence disclosure within the CFT field represents a professional challenge, in particular since CFT treatments are found to be effective treatments for couple violence (Stith, McCollum, Amanor-Boadu, & Smith, 2012).

To help clinicians detect cases where violence is part of the problem, even if it is not explicitly mentioned in the referral or by the clients, therapists need knowledge of theory, typical patterns of partner violence (see. e.g., Bensimon & Ronel, 2012; Entilli & Cipolletta, 2016; Walker, 1979) and predictors of couple and family violence. Knowledge of predictors of couple and family violence can also help therapists understand the complexity of this type of violence and thereby provide more appropriate therapy. Moreover, developers of treatment for couple and family violence can use the predictors to tailor new and possibly more efficient treatments.

Even if predictor studies of couple and family violence are rare on a clinical CFT sample, predictor studies of couple violence have been conducted on other samples (see e.g., K. O'Leary & Woodin, 2009, for a review). Unlike many other health problems, few social and demographic characteristics define risk groups for intimate partner violence (Jewkes, 2002). Nevertheless, couple violence relates to other personal and relational factors. Low income and low level of education have been found to be associated with higher prevalence of couple violence (see. e.g. Pan, Neidig, & O'leary, 1994; Stith et al., 2011). Furthermore, it has been determined that physical couple violence is associated with alcohol and substance abuse, both at present time and in adults' family of origin (e.g., Coker, Smith, McKeown, & King, 2000; Fals-Stewart, 2003; Fals-Stewart, O'Farrell, Birchler, Córdova, & Kelley,

2005; Stith et al., 2012). In addition, the use of violence in close relations is confirmed to be closely related to anger (Simpson, Doss, Wheeler, & Christensen, 2007).

A close relationship between physical and sexual abuse has been found in several studies (e.g. Simpson, Atkins, Gattis, & Christensen, 2008; Simpson et al., 2007; Stith et al., 2012). However, we could not find any study that investigated the relationship between sexual satisfaction and couple violence. Issues of sexual satisfaction are common in CFT treatment. Since good sexual satisfaction is found to be associated with relationship satisfaction, love, commitment, and stability (Butzer & Campbell, 2008; Sprecher, 2002; Young, Luquis, Denny, & Young, 1998) it is reasonable to believe that there is less couple violence when the sexual satisfaction is high. Most couples argue and fight over practical issues like household chores, and it is common that some of these issues occur repeatedly (Gottman & Silver, 2015). These authors argue that the underlying issues are often values or expectations: e.g., if the expectation of one partner about the other partner's household chores is higher than what the other expects of herself, there is likelihood of conflict. This could lead to tiring arguments with high tension or even violent outcomes. However, different people have different conflict management styles that have been assessed in relation to couple violence (see e.g., Gottman & Silver, 2015; Simpson et al., 2007; Vivian & Malone, 1997), and the findings suggest some styles of conflict management to be far more associated with violence than others. The authors define these in different ways, but there is a consensus that self-control is viewed as a positive asset in high conflict.

As mentioned, the knowledge of couple and family violence within a clinical CFT sample is limited. There are even fewer studies of predictors of couple and family violence within this group of clients, and no studies where family violence has been included. In this study, we investigated some characteristics among CFT clients with defined issues of violence. Furthermore, we investigated a model that may be helpful in discovering couple and family violence, thus helping therapists to assess what actions to take in therapy to prevent further physical violence. To our knowledge, this is the first study to investigate couple violence on a clinical CFT sample outside the US, and the first study to investigate predictors of both physical couple violence and physical family violence in a CFT sample.

The research questions are:

1. What is the prevalence of physical couple and family violence in a clinical sample?
2. What identifies clients who experience physical couple violence?
3. What predicts physical couple and family violence?

1 Material and Methods

We have conducted a study of clients seeking CFT treatment by using quantitative data collected with the Systemic Therapy Inventory of Change (STIC) (Pinsof et al., 2009) feedback system. STIC measures several aspects of clients' lives including physical couple and family violence together with several items that possibly can predict this violence.

1.1 Sample

The initial sample for this study consisted of 830 clients above 18 years of age (mean age 40.3 years (SD = 8.5); age range 18 to 72 years; 51.8 % were women, mean 2.3 children (SD = 1.1), see table in the supplementary material for more details)[1]. Data collection started in March 2010 and was ended in April 2016 and the sample consists of data from all three levels in the stepped level of care within CFT in Norway. The first and second levels of care were represented by outpatient agencies. At the first level, no referral was needed. The third and final level of care was represented with an inpatient CFT agency, where a referral was needed. In all, 56% (462) outpatients and 54% (368) Inpatients were included in the study. All participants completed the Systemic Therapy Inventory of Change (STIC) (Pinsof et al., 2009) initial questionnaire in Norway during a pilot (Tilden, Håland, Hunnes, Fossli, & Oanes, 2015) and an RCT study (Tilden et al., In press).

1.2 Ethics

Written informed consent for collecting the project data was obtained from each participant. This study was approved by the Modum Bad Ombudsman for Data Protection and the Regional Ethics Committee for Medical Research with human subjects (2017/96/REK sør-øst C). The primary study is registered at ClinicalTrials.gov. Since the data originates from regular clinical practice, no inclusion or exclusion criteria have been used except for the ones each site has for accepting clients for treatment.

1.3 Measurements

Measures. Systemic Therapy Inventory of Change (STIC – Pinsof et al., 2009) is an assessment and feedback system in which clients fill out a questionnaire before each therapy session. Via electronic devices, the clients evaluate their response to treatment, including progression, and alliance to the therapist (Pinsof et al., 2009; Tilden et al., 2010). Client evaluations are processed into a report that is returned to the therapists who can use this information as the basis for understanding and hypotheses in the clinical assessment of the current client. The response options in STIC are on a 5-level scale from worst to best option. Modules are added depending on the therapy mode and number of clients in the therapy system. The questions cover six subscales, individual problems and strengths, family of origin, relationship with partner, family/household, children's problems, and strength and relationship with child or children. On average, it takes 45 minutes to complete the STIC questions before the first treatment session. Before every subsequent session, the clients complete a short version of STIC that takes 7-8 minutes to fill out. Because this study is cross-sectional by making use of data before the first session only, the intersession data were not included. STIC has good internal reliability (Pinsof et al., 2009; Zahl-Olsen, Gausel, Tilden, & Håland, In review; Zinbarg et al., 2018) with a Cronbach's alpha as high as 0.94 for the different subscales.

The STIC consists of several subscales that further contain factors and items, and some of these address the topic for this study. The response options to those questions were, not at all/never, rarely, sometimes, often, and all of the time. The Relationship with partner (RWP) scale has one item addressing physical violence between the members of the couple: "We get into shoving or hitting each other when we fight". The family household (FH) scale has two items addressing physical violence exerted within the family: "Someone in my family is physically abusive to other family members" (item 1), and "There is someone in my family who pushes other family members around physically to get his or her way" (item 2). We combined these two items to one family violence item

[1] 813 defined themselves as heterosexual, 8 as bisexual and 2 as homosexual/lesbian.

in our analysis. We hypothesized that both the couple violence item and the family violence item would load from the same latent variable Physical Violence. Further, the individual problems and strength (IPS) scale, the relationship with partner (RWP) scale, and the family of origin (FOO) scale all have variables we, based on the presented theory and the literature review above, hypothesize are predictors for the underlying latent variable – Physical Violence. We also modeled that perceived anger towards the partner and level of expectation of household chores were predictors of Physical Violence. We hypothesized two negative predictors of Physical Violence: Self-control of thoughts, impulses and rage as the first one, and sexual satisfaction within the relationship as the second. See table 1 for a full list of items included in the variables[2].

Insert table 1, List of items used to model …, about here

1.4 Statistical analysis

We used IBM SPSS Statistics 25 for descriptive, bivariate and multivariate analysis, and Amos 25 for Structural Equation Modelling (SEM) analysis. Descriptive, correlation, t-tests and crosstab analysis were performed as instructed by Field (2018) to describe the sample, calculate the statistics and to compare the group of clients who indicated couple violence with the clients who did not. Further, the statistical analyses included two Multivariate General Linear Models, which were performed as described by Field (2018), where existence of couple violence were set as the fixed factor.

In this study, we used Structural Equation Modelling (SEM) to predict Physical Violence. This method has several advantages compared to more conventional statistical techniques (Kline, 2016). For example, in multiple regression, it is an assumption that all predictors are measured without errors. This is routinely violated in practice, but when using SEM, we can make explicit representations of measurement errors. SEM also makes it possible to model the correlations between the variables and include that as part of the analysis. Finally, SEM involves significance testing of whole statistical models and not just individual effects (Kline, 2016). Based on our expectations, we modeled the latent variable Physical Violence, which is expressed as physical violence between the couple and physical violence between others in the family. Further, we hypothesized that predictors in the model were the partner's expectations towards each other, levels of anger, sexual satisfaction and self-control, as described below.

2 Results

2.1 Descriptive analyses.

Our first approach to the research question addressing the prevalence of couple and family violence was to analyze the clients' responses to the item "We get into shoving or hitting each other when we

[2] Since prior studies (see e.g., Coker et al., 2000; Fals-Stewart, 2003; Stith et al., 2012) indicated that alcohol abuse might be of importance and STIC provided measures of alcohol abuse at three different levels: for themselves, for their partner and in their family of origin, we wanted to assess if these were significant contributors to the model also in our clinical sample.

fight". As many as 20.4% (n = 169) confirmed that this described their relationship from "rarely to all of the time". This group is from now on called the Couple Violence group (CV) as discriminant from the rest of the group labeled No Couple Violence group (NCV). Within the CV group 84% (142) reported rarely, 13% (22) sometimes, 3% (5) often, and 0% (0) all the time. Our next approach to the first research question was to analyze the clients' responses to family violence. As many as 24.9% (207) responded from "rarely to all of the time".

Our second research question addressing the identification of clients who experience physical couple violence was assessed by comparing the CV and the NCV groups. We found that the CV group had significantly more prior experience with therapy (χ^2 (3) = 8.165, p =.043) and had lower income (χ^2 (2) = 13.612, p =.001) compared to the NCV group. However, the two groups did not differ on measures of age and education. An important difference between the two groups was the presence of family violence. In the CV group, as many as 49.7% reported that family violence was present. This was significantly higher (χ^2 (1) = 84,324, p < .001) than in the NCV group reporting 18.0%. The CV group was significantly more distressed than the NCV group, on all the measures included in the model. Using Roy's Largest Root, there was a significant effect of experiencing couple violence on couple and family violence, Θ = 6.34, F (2,691) = 2191.40, p < .001. For the violence items higher values are better and for family violence the sample of clients in the CV sample had a mean of 4.46 (SD = .61) and the NCV group 4.77 (SD = .55). For couple violence the sample of clients in the CV sample had a mean of 3.81 (SD = .46) and the NCV group 5.00 (SD = .00). Further, using Roy's Largest Root, there was a significant effect of experiencing couple violence on the predictors in the model, Θ = 0.19, F (4,762) = 35.33, p < .001. For Anger and Expectation higher values are better and the mean values on Anger was 3.20 (SD=.89) for the CV sample and 3.82 (SD = .97) for the NCV sample. The mean values on Expectation was 2.89 (SD=.81) for the CV sample and 3.56 (SD = .77) for the NCV sample. For Sexual satisfaction and self-control lower values are better and the mean values for Sexual satisfaction was 2.87 (SD=.99) for the CV sample and 2.66 (SD = 1.08) for the NCV sample on. The mean values for the CV sample was 1.96 (SD=.83) and 1.63 (SD = .71) for the NCV sample on Self-control. See table in the supplementary material for further information.

2.2 Model analysis: Structural Equation Modeling (SEM)

Our third research question addressed possible predictors of couple and family violence in the total sample. The model was constructed with the variables as previously described[3], with four predictors and one latent variable, Physical Violence, as the outcome variable. The model as displayed in figure 1, explained 53 % of the change in Physical Violence and demonstrated very good fit for this sample. The chi-square was: χ^2 (27) = 78.672, p =.000, which indicates that the model fits the data well. However, given the complexity of the model, chi-square is an inadequate test of model fit (Kline, 2016). A better way to test how well our hypothesized model fit the data is provided by a χ2/df ratio below 3 (2.914) and several chi-square-based fit indices above .900 (incremental fit index [IFI] = .967, comparative fit index [CFI] = .967). In, addition, good model fit was shown by our observation

[3] Results from the CFA on the model was χ2 (15) = 64.211, p <.001. χ2/df ratio = 4.281, IFI = .964, CFI = .964, RMSEA = .063. The correlations were: .44 between Expectation and Anger, -.55 between Expectation and Sexual satisfaction, -.15 between Expectation and Self-control, -.41 between Anger and Sexual satisfaction, -.39 between Anger and Self-control, and .12 between Sexual satisfaction and Self-control. The factor loadings from Expectation was .80 to Expect 1, .47 to Expect 2 and .56 to Expect 3. The factor loadings from Sexual satisfaction was .88 to Sexsat 1, .77 to Sexsat 2. The factor loadings from Self-control was .58 to Selfcont 1 and .75 to Selfcont 2.

of a residual index, where lower is better (root-mean-square error of approximation [RMSEA] = .048 with the confidence intervals, .036 - .061; see Kline, 2016). The model was stable also when controlling for demographics[4]. The standardized solution is shown in Figure 1[5], and the scale inter-correlations and descriptive statistics is presented in table 2.

Figure title: Figure 1. Predictive model of Physical Violence.

Insert figure 1, SEM model, about here

Note: All covariance's and factor loadings were statistically significant ($p < .05$).

Insert table 2, scale inter-correlations and descriptive statistics, about here

The two strongest predictors of Physical Violence were Expectation ($\beta = .73$, $p < .001$) and contrary to our expectation, Sexual satisfaction ($\beta = .42$, $p < .001$). Anger was also a positive predictor ($\beta = .13$, $p < .001$). Self-control was a statistically significant negative predictor of Physical Violence ($\beta = -.21$, $p = .004$). The correlation analysis showed that all the predictors were statistically significant related to each other (all *p's* < .05). The medium levels of correlation and the variation of the correlation coefficients indicate that the included variables measure different aspects of the client's life.

3 Discussion

We found that 20.4 % of the clients reported that physical violence occurs in their relationship. Furthermore, as many as 49.7 % of the clients reporting physical couple violence also reported physical family violence. The clients with physical couple violence had lower income and more prior experience with therapy. The strongest predictor of Physical Violence were Expectation, while Self-control was a negative predictor.

3.1 Prevalence of physical couple violence in a clinical sample

The prevalence of couple violence in the clinical CFT sample in the current study is 20.4 % which is more than four times as high as the prevalence of couple violence in the general population in Norway (Haaland et al., 2005). More precisely, this study identifies physical violence and not a more generally defined violence. Furthermore, in this study, we have specified the questions with regard to the current relationship. This is contrary to most population studies, which ask if people have

[4] None of the three alcohol abuse items improved the model.

[5] Results from the analyses of the same model using the CV sample (N = 165) were $R^2 = .12$, $\chi2$ (27) = 37.476, p =.087. $\chi2$/df ratio = 1.388, IFI = .969, CFI = .966, RMSEA = .048. Analyzing the model on the NCV sample (N = 604) requires removal of the latent variable Physical Violence since Couple Violence is zero for all participants in this sample. Thus, the outcome variable is the manifest variable Family Violence. The results of this model were, $R^2 = .04$, $\chi2$ (19) = 63.127, p <.001. $\chi2$/df ratio = 3.322, IFI = .956, CFI = .955, RMSEA = .062.

experienced couple violence in *any* intimate relationship until now. US studies of couples seeking therapy have found frequencies of couple violence up to 61% (Cascardi et al., 1992; Jouriles & O'leary, 1985; K. D. O'Leary, Vivian, & Malone, 1992; Simpson et al., 2007; Stith et al., 2011; Vivian & Malone, 1997). However, in these studies, they have used a more general definition of couple violence that includes psychological, physical and sexual violence without differentiating between these. Hence, more research is needed where the types of violence are accounted for.

3.2 **Couple violence and family violence occurs simultaneously**

Our descriptive analyses found that there is a significantly higher distribution of family violence in the CV group than in the NCV group. This co-occurrence of violence in the couple and violence in the family is in line with the findings of Slep and O'leary (2005) and Apple and Holden (1998). However, the high prevalence of both couple and family violence among clients seeking CFT identified in our study indicates that every therapist meets clients with these issues frequently, even if it is not addressed. Further efforts should therefore be made in CFT to discover and end couple and family violence than is the case currently. Since some research indicates that clients are more likely to reveal family violence when responding to a questionnaire compared to when questioned face to face by a therapist, e.g. (Andersen & Svensson, 2013; Ehrensaft & Vivian, 1996; Zahl-Olsen & Oanes, 2017) we suggest implementing systematic screening for both couple and family violence, for instance as part of a feedback system.

Family violence co-occurs with couple violence for several reasons. Edelson (1999) found that children were involved in the violence among the adults, while Raundalen (2004) found that children tried to intervene to stop the violence. Further, parents who conduct physical violence against their partner also do so towards the children in the family (Appel & Holden, 1998; Slep & O'leary, 2005). It is found that children exposed to severe anger and aggression in their domestic environment increase their risk of becoming more aggressive (Raundalen, 2009). In addition, children exposed to violence in their home have higher risks of experiencing violence with others later in life compared to those who did not encounter such violence in their childhood home (see, e.g., Wolf & Foshee, 2003; Øverlien, 2012). Indeed, children exposed to violence in their childhood also have a greater risk of being exposed to adult violence (Renner & Slack, 2006) later in life. Thus, there is a high risk that violence is transferred from one generation to the next. Finally, these findings support the theoretical argumentation of Gausel and Leach (2011). In their theoretical framework, children who are exposed to violence in their home might appraise this violation as a sign of threat to their worth, and by such, a sign of disrespect. However, it can also be appraised as a global defect in their family, i.e., we are violent people, which again might provoke anger. Combined, these two ways to appraise violence might encourage further violence to end the ongoing disrespect, a last effort to protect their social image as someone who is worthy of respect (Gausel, 2013; Gausel et al., 2018).

3.3 **Differences between the samples with and without couple violence**

In our study, the prevalence of couple and family violence is higher among those with low income. This is a finding similar to other studies, see e.g., Gelles (1997), Rennison & Planty (2003) and Weede (1981). Further, it is an interesting finding that the CV group has received more prior treatment than the NCV group. However, this might indicate that the CV group consists of families with more severe issues than the NCV group and therefore needs more treatment. If so, it is an

empirical question whether the treatment they received was helpful if the violence was addressed as a topic in therapy, or whether – as previously mentioned – the therapeutic interventions so far did not reveal the ongoing violence. There were significant differences between the CV sample and the NCV sample on the levels of distress on the predictor variables. However, the differences were small and both groups indicated distress at all four predictors. Therefore, we continued to investigate the individual differences on the total sample.

3.4 **Expectation and anger**

Our model demonstrated that the experiences of unreasonable expectation from one partner towards the other in relation to household chores was the strongest predictor of Physical Violence. In other words, the more experiences of unreasonable expectation from the partner, the more physical violence in the relationship. This is in agreement with research finding that marital discord serves as a strong predictor of both mild and severe husband-to-wife physical aggression (Capaldi et al., 2012; Pan et al., 1994; Stith et al., 2008). However, since males and females are found to be equally exposed to couple violence (Haaland et al., 2005; Jose & O'Leary, 2009; Stith et al., 2011; Thoresen & Hjemdal, 2014) our study did not differentiate between husband-to-wife and wife-to-husband aggression.

When viewing this strong prediction from Expectation to Physical Violence from the theoretical perspective of Gausel and Leach (2011) we understand why this link is so strong. When someone experience that the other person in their relationship is viewing them as someone who does not manage to live up to what is expected, they might feel inferior, and thus wanting to act, either to withdraw, which is hard in a close relationship, or by using verbal or physical force.

We found that expectation and anger were significantly correlated. It makes sense that when a partner experiences herself or himself to be criticized and treated unfairly, he or she experiences anger (Isdal, 2000). The other way around also makes sense. Gottman and Silver (2015) describe how a person with anger and arousal interprets even neutral words as negative. This is also in line with the conceptual theory of (Gausel & Leach, 2011) that being criticized can lead to a feeling of inferiority and rejection which again is associated to anger. However, anger in our model is the third strongest predictor of Physical Violence. Apart from cases where dominating the other partner is the issue it is hard to think of couple violence occurring without anger, but it is important to point out that experiencing the feeling of anger does not say it has to lead to physical couple violence (Leone, Johnson, & Cohan, 2007). Leone and colleagues (2007) argue that violence is usually about lack of affective regulation, communication difficulties, and the lack of skills for problem-solving. Thus, if a person has not learned to be conscious of and learned how to handle basic feelings such as anger, this can, in turn, lead to unwanted behavior like physical violence.

3.5 **Sexual satisfaction**

Among our two expected negative predictors for Physical Violence, better Self-control, and higher Sexual satisfaction, only the first was confirmed. Moreover, our SEM analyses indicated that Sexual satisfaction in the relationship predicted Physical Violence. This is contrary to our expectations and contrary to studies of sexual satisfaction elsewhere in the field of CFT where sexual satisfaction is found to be associated with good and healthy relationships (see e.g., Butzer & Campbell, 2008;

Litzinger & Gordon, 2005; Sprecher, 2002; Yeh, Lorenz, Wickrama, Conger, & Elder Jr, 2006). Besides, Litzinger & Gordon (2005) found that sexual satisfaction partially compensates for the negative effects of poor communication. However, one possible way to understand our finding that higher sexual satisfaction predicts couple violence is that couples who emotionally live distant from each other have very little contact and do not argue and fight. In consequence, they do not have much sexual intimacy either. This group might level out those who occasionally fight and have a satisfactory sexual life. This inference is supported by the fact that 84% (142) of those indicating physical couple violence answered that the physical violence occurred seldom. Furthermore, we found a normal distribution of sexual satisfaction within the group who indicated that they seldom experienced couple violence. Walker's (1979) cycle of violence identifies a *honeymoon phase* that follows a phase of acute battering. In this honeymoon phase, identified as calm and loving, satisfying sex can be present. This could explain why sexual satisfaction came out as a predictor of physical violence in this study. However, we have not been able to find research on the direct connection between sexual satisfaction and violence in relationships. Hence, we suggest further research on the relationship between sexual satisfaction and couple violence. Finally, the conceptual model of shame (Gausel & Leach, 2011; Gausel, 2013; Gausel et al., 2018) explains that if the moral failure is appraised as a self-defect that is mendable the person will try to repair what is broken. Sex could be a way to strengthen and restore one's self-image as someone to love and respect, and even more to build stronger social bonds to the partner and thus prevent condemnation.

3.6 Self-control

As mentioned, self-control was found to be a negative predictor of Physical Violence. Thus, the more self-control the clients has the lesser physical violence they experience. Furthermore, this seems to be a variable that is not closely related to the other variables in our model. In other words, this may be a variable that is stable even when the others vary. This seems especially true in relation to sexual satisfaction and expectation. Thus, from a clinical point of view, strengthening the clients' self-control could contribute to bringing stability into the relationship and decreasing physical violence.

4 Strengths and limitations

It is a strength of our study that we have data from all three different levels of CFT treatment in Norway, including samples from low threshold outpatient clinics without the need of a referral, an outpatient clinic where referral is needed, to an inpatient clinic where referral is needed. However, that the sample, thereby, is heterogeneous could also be considered a limitation because we have not analyzed the differences between the clinics. A second strength is the size of the data set and furthermore that approximately half of the data stems from an RCT with strict control of data collection. A third strength is that the type of violence is specified as physical violence in the current relationship. However, it is a limitation that the clients did not report who exercised the couple violence: whether it is themselves as perpetrators, whether they are exposed to it by their partner, or whether it is both. A second limitation is that we could not differentiate between severe physical violence and minor physical violence. However, one could argue that by dividing the sample into three groups rather than two we could gain more information: 1= no violence, 2= seldom violence, 3= often or more. However, we had two reasons for not doing so. First, even if violence seldom occurs, we do not know how severe it is, and a few very severe incidents might be as powerful as

many less severe. Second, if violence only rarely occurs it can still be an important part of their family life. A third limitation is the narrow focus on anger as the only emotion identified as predictor of family violence in our model. Undoubtedly, there are other emotions, such as a feeling of rejection (Gausel, 2013; Leary, Twenge, & Quinlivan, 2006) or of hate (Staub, 2005), that predict violence and anti-social behavior. We acknowledge that we are not able to identify whether one or both parties are conducting violence. However, recent research on violent conflicts suggest that in most violent conflicts both parties exercise violence but prefer to construe themselves as victims instead of perpetrators (e.g., Gausel, Leach, Mazziotta & Feuchte, 2018; Mazziotta, Feuchte, Gausel & Nadler, 2014). By such, it's not unthinkable that if we had been able to explore who exercised violence in our study parties would tend to see themselves as victims and the other as the perpetrator - therefore skewing the reports of violence towards the other instead of reporting their own perpetration (for a discussion, see; Gausel et al., 2018). Some might question if the sample of clients who experienced physical violence is representative for this group, since 84 % of them indicate that the physical violence is experienced rarely and 13 % sometimes. We share this concern and want to make clear that this is how the clients have reported physical violence in their current relationship, and that under-reporting is possible. However, this sample might not be representative for clients e.g. seeking therapy specialized at domestic violence or coming to women's shelters.

5 Conclusion and clinical implications

This study has found that the prevalence of couple and family violence in a clinical sample is high, indicating that many CFT therapists encounter this topic in therapy. Because this topic may remain undisclosed during treatment, we assume that a higher portion of CFT therapists may relate to this topic indirectly because when this remains a secret, it may impact a great deal on the presented topics in therapy. The notion that men cannot be victims of physical and psychological violence by their partners (see. e.g., Entilli & Cipolletta, 2016) may impose a considerable barrier for therapists to interpret such signs. Since the immediate and consequential damages of couple and family violence are grave for the adult, their children and their future generations, therapists need tools to discover these issues. Knowledge of the predictors in this study can possibly help to uncover physical abuse in couple relationships and families. Therapists that have little knowledge of physical violence and predictors of such violence risk to miss the opportunity to assess signs of an abusive relationship, or to seek further assistance from child protection services and police. Knowledge of predictors of physical violence can also bring the confidence in the CFT therapists to dare to ask more specific and handle the case in a proper way. CFT therapists are in contact with more couples and families where violence occur, than the police (see e.g., SSB, 2016 and Reigstad et al., 2006) and are enforced to notify when this is uncovered (Barnevernloven 1993). In Norway, this law also enforce teachers, social workers and therapists to be aware of circumstances that are harmful (Driscoll, 2018). This challenges CFT therapists, especially those inspired by social constructionist theories which not want to enter areas in people's lives that the clients do not directly address (Anderson, 2001). When it comes to violence we cannot allow therapists to be more influenced by their own perspectives than the data about the situation (Jensen, 2008). It is pivotal to enhance the knowledge of predictors of physical violence that could be noticed by CFT therapists and allow a more effective screening.

In addition, as shown in our model, we have addressed certain predictors that can help to discover physical violence in ordinary CFT practice. Furthermore, our findings could assist therapists in assessing what actions to take in therapy to prevent further physical violence. For example, by focusing the therapy on expectations towards each other's participation in household duties and what lays behind those expectations. It is possible that, through therapy, focusing on attaining mutual understanding of the expectation towards each other this tension will go down and thus prevent physical violence. Further, as the conceptual theory of Gausel and Leach (2011, Gausel et al., 2018) explains, the feeling of rejection, through the experiences of loss of social image (others' view of self), and the feeling of inferiority can lead to violent acts. Hence, if the focus in therapy is on understanding and acceptation of the partners' view of self, this will probably impact on reduced uncertainty about the partner's view of self. Since most partners want to stay in the relationship even if violence is exerted (Stith et al. 2011) it is reasonable to think that their view of their partners is mostly positive. Reduced uncertainty implies fewer chances of viewing self-defects as global, instead increasing the chance that failures are specific and thus can be more easily dealt with. In return, as Gausel and Leach (2011) argue, this should lead to acts of reconciliation and amendments.

These findings address clinical implications on how to treat couples and families where abuse is a topic. However, it is outside the scope of this paper to give a thorough clinical guide on how to do so. There is a rich literature that the reader is encouraged to search for on this topic.

6 Suggestions for further research

We suggest further research on the prevalence and predictors of different types of couple and family violence (verbally, sexually and physically) in clinical samples. Based on our unexpected finding of the relationship between sexual satisfaction and physical violence we suggest investigating this relationship in further research. Our study has also found that alcohol abuse was a less important predictor of physical violence than the literature suggests. Thus, we encourage further research on aspects of expectation, anger, and self-control as predictors for physical couple and family violence.

7 References

Amnesty International. (2013). Fakta om vold mot kvinner. Retrieved from http://www.amnesty.no/aktuelt/flere-nyheter/arkiv-bakgrunn/fakta-om-vold-mot-kvinner

Andersen, A., & Svensson, T. (2013). Internet-based Mental Health Services in Norway and Sweden: Characteristics and Consequences. *Administration and Policy in Mental Health and Mental Health Services Research, 40*(2), 145-153. doi:10.1007/s10488-011-0388-2

Appel, A. E., & Holden, G. W. (1998). The co-occurrence of spouse and physical child abuse: A review and appraisal. *Journal of Family Psychology, 12*(4), 578.

Barnevernloven. (1993). *Lov om barneverntjenester*.

Bensimon, M., & Ronel, N. (2012). The flywheel effect of intimate partner violence: A victim–perpetrator interactive spin. *Aggression and Violent Behavior*, *17*(5), 423-429.

Butzer, B., & Campbell, L. (2008). Adult attachment, sexual satisfaction, and relationship satisfaction: A study of married couples. *Personal relationships, 15*(1), 141-154.

Cascardi, M., Langhinrichsen, J., & Vivian, D. (1992). Marital aggression: Impact, injury, and health correlates for husbands and wives. *Archives of internal medicine, 152*(6), 1178-1184.

Coker, A. L., Davis, K. E., Arias, I., Desai, S., Sanderson, M., Brandt, H. M., & Smith, P. H. (2002). Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine, 23*(4), 260-268.

Coker, A. L., Smith, P. H., McKeown, R. E., & King, M. J. (2000). Frequency and correlates of intimate partner violence by type: physical, sexual, and psychological battering. *American journal of public health, 90*(4), 553.

Driscoll, H. (2018). Taushetsplikten i arbeid med barn og unge. In B. Johannessen & T. Skotheim (Eds.), *Barn og unge i midten. Tverrfaglig og tverretatlig arbeid i barn og unges oppvekst* (pp. 148-172). Oslo: Gyldendal akademisk.

Edleson, J. L. (1999). Children Witnessing adult domestic violence. *Journal of Interpersonal Violence, 14*(8), 839-870.

Ehrensaft, M. K., & Vivian, D. (1996). Spouses' reasons for not reporting existing marital aggression as a marital problem. *Journal of Family Psychology, 10*(4), 443.

Fals-Stewart, W. (2003). The occurrence of partner physical aggression on days of alcohol consumption: A longitudinal diary study. *Journal of Consulting and Clinical Psychology, 71*(1), 41.

Fals-Stewart, W., O'Farrell, T. J., Birchler, G. R., Córdova, J., & Kelley, M. L. (2005). Behavioral couples therapy for alcoholism and drug abuse: Where we've been, where we are, and where we're going. *Journal of Cognitive Psychotherapy, 19*(3), 229-246.

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. London, Sage Publishing

Fusco, R. A., & Fantuzzo, J. W. (2009). Domestic violence crimes and children: A population-based investigation of direct sensory exposure and the nature of involvement. *Children and Youth Services Review, 31*(2), 249-256.

Gausel, N. (2013). Self-reform or self-defense? Understanding how people cope with their moral failures by understanding how they appraise and feel about their moral failures. In M. Moshe & N. Corbu (Eds.), *Walk of shame* (pp. 191-208). NY, USA: Hauppauge.

Gausel, N., Leach, C. W. (2011). Concern for self-image and social image in the management of moral failure: Rethinking shame. *European Journal of Social Psychology,41, 468-478.* doi: 10.1002/ejsp.803

Gausel, N., Leach, C. W. Mazziotta, A., & Feuchte, F. (2018). Seeking revenge or seeking reconciliation? How concern for social-image and felt shame helps explain responses in reciprocal intergroup conflict. *European Journal of Social Psychology, 48*, 62-72. doi: 10.1002/ejsp.2295

Gelles, R. J. (1997). *Intimate violence in families*: Sage.

Golding, J. M. (1999). Intimate partner violence as a risk factor for mental disorders: A meta-analysis. *Journal of Family Violence, 14*(2), 99-132.

Gottman, J. M., & Silver, N. (2015). *The seven principles for making marriage work: A practical guide from the country's foremost relationship expert*: Harmony.

Haaland, T., Clausen, S.-E., & Schei, B. (2005). *Vold i parforhold - ulike perspektiver : resultater fra den første landsdekkende undersøkelsen i Norge* (Vol. 2005:3). Oslo: Norsk institutt for by- og regionforskning.

Isdal, P. (2000). *Meningen med volden*. Oslo: Kommuneforlaget.

Jensen, P. (2008). The Narratives which connect...: a qualitative research approach to the Narratives which connect terapists' personal and private lives to their family theraphy practices. *Doctoral thesis at University of East London in conjunction with the Travistock Clinic.*

Jewkes, R. (2002). Intimate partner violence: causes and prevention. *The Lancet, 359*(9315), 1423-1429.

Jose, A., & O'Leary, K. D. (2009). Prevalence of partner aggression in representative and clinic samples.

Jouriles, E. N., & O'leary, K. D. (1985). Interspousal reliability of reports of marital violence. *Journal of Consulting and Clinical Psychology, 53*(3), 419.

Kimball, E. (2016). Edleson Revisited: Reviewing Children's Witnessing of Domestic Violence 15 Years Later. *Journal of Family Violence, 31*(5), 625-637. doi:10.1007/s10896-015-9786-7

Kitzmann, K. M., Gaylord, N. K., Holt, A. R., & Kenny, E. D. (2003). Child witnesses to domestic violence. A meta-analytic review. *Journal of Consulting and Clinical Psychology, 71*(2), 339-352.

Kline, R. B. (2016). *Principles and Practice of Structural Equation ModelingFourth Edition* (Fourth edition ed.). New York: The Guilford Press.

Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). *World report on violence and health*. Geneva: World Health Organization.

Leary, M. R., Twenge, J. M., & Quinlivan, E. (2006). Interpersonal rejection as a determinant of anger and aggression. *Personality and Social Psychology Review*(10(2)), 111-132.

Leone, J. M., Johnson, M. P., & Cohan, C. L. (2007). Victim help seeking: Differences between intimate terrorism and situational couple violence. *Family Relations, 56*(5), 427-439.

Litzinger, S., & Gordon, K. C. (2005). Exploring relationships among communication, sexual satisfaction, and marital satisfaction. *Journal of Sex & Marital Therapy, 31*(5), 409-424.

Loring, M. T. (1994). Emotional abuse. New York, NY, US: Lexington Books/Macmillan.

Mazziotta, A., Feuchte, F., Gausel, N**.** & Nadler, A. (2014). Does remembering past ingroup harmdoing promote postwar cross-group contact? Insights from a field-experiment in Liberia. *European Journal of Social Psychology, 44,* 43-52. doi: 10.1002/ejsp.1986

Middelborg, J., & Samoilow, D. K. (2014). *Tryggere barndom : et behandlingsperspektiv på vold i familien*. Oslo: Gyldendal akademisk.

NKPH, N. k. f. p. h. (2014). Vold og overgrep i nære relasjoner. Retrieved from http://www.napha.no/content/13928/Vold-og-overgrep-i-nare-relasjoner

O'Leary, K., & Woodin, E. M. (2009). *Psychological and physical aggression in couples: Causes and interventions*: American Psychological Association.

O'Leary, K. D., Vivian, D., & Malone, J. (1992). Assessment of physical aggression against women in marriage: The need for multimodal assessment. *Behavioral Assessment*.

Ormhaug, S. M., Jensen, T. K., Hukkelberg, S. S., & Egeland, K. (2012). Traumer hos barn - blir de gjemt eller glemt? ; kartlegging av traumatiske erfaringer hos barn og unge henvist til BUP. *Tidsskrift for Norsk psykologforening, 49*(3), 235-240.

Pan, H. S., Neidig, P. H., & O'leary, K. D. (1994). Predicting mild and severe husband-to-wife physical aggression. *Journal of Consulting and Clinical Psychology, 62*(5), 975.

Pinsof, W. M., Zinbarg, R. E., Lebow, J., Knobloch-Fedders, L. M., Durbin, E., Chambers, A., . . . Friedman, G. (2009). Laying the foundation for progress research in family, couple, and individual therapy: The development and psychometric features of the initial Systemic Therapy Inventory of Change. *Psychotherapy Research, 19*(2), 143-156.

Raundalen, M. (2004, 15.4). Barn som lever med vold i familien *Bulleteng,* pp. 6-16.

Raundalen, M. (2009). Vold mot barn - skader. *Bulleteng nr. 6,* pp. 2-32.

Reigstad, B., Jørgensen, K., & Wichstrøm, L. (2006). Diagnosed and self-reported childhood abuse in national and regional samples of child and adolescent psychiatric patients: Prevalences and correlates. *Nordic Journal of Psychiatry, 60*(1), 58-66. doi:10.1080/08039480500504933

Renner, L. M., & Slack, K. S. (2006). Intimate partner violence and child maltreatment: Understanding intra-and intergenerational connections. *Child abuse & neglect, 30*(6), 599-617.

Rennison, C., & Planty, M. (2003). Nonlethal intimate partner violence: Examining race, gender, and income patterns. *Violence and Victims, 18*(4), 433.

Schacht, R. L., Dimidjian, S., George, W. H., & Berns, S. B. (2009). Domestic violence assessment procedures among couple therapists. *Journal of Marital and Family Therapy, 35*(1), 47-59.

Simpson, L. E., Atkins, D. C., Gattis, K. S., & Christensen, A. (2008). Low-level relationship aggression and couple therapy outcomes. *Journal of Family Psychology, 22*(1), 102-111.

Simpson, L. E., Doss, B. D., Wheeler, J., & Christensen, A. (2007). Relationship violence among couples seeking therapy: Common couple violence or battering? *Journal of Marital and Family Therapy, 33*(2), 270-283.

Slep, A. M. S., & O'leary, S. G. (2005). Parent and partner violence in families with young children: rates, patterns, and connections. *Journal of Consulting and Clinical Psychology, 73*(3), 435.

Sprecher, S. (2002). Sexual satisfaction in premarital relationships: Associations with satisfaction, love, commitment, and stability. *Journal of sex research, 39*(3), 190-196.

SSB. (2016). Familievern, 2014. Retrieved from http://www.ssb.no/sosiale-forhold-og-kriminalitet/statistikker/famvern/aar/2015-10-27?fane=tabell&sort=nummer&tabell=243787

Stacey, W. A., Hazlewood, L. R., & Shupe, A. D. (1994). *The violent couple*. Westport, CT: Praeger.

Staub, E. (2005). The Origins and Evolution of Hate, With Notes on Prevention. In R. J. Sternberg (Ed.), *The psychology of hate* (pp. 51-66). Washington, DC, US: American Psychological Association.Stith, S. M., McCollum, E. E., Amanor-Boadu, Y., & Smith, D. (2012). Systemic perspectives on intimate partner violence treatment. *Journal of Marital and Family Therapy, 38*(1), 220-240.

Stith, S. M., McCollum, E. E., & Rosen, K. H. (2011). *Couples therapy for domestic violence: Finding safe solutions.* American Psychological Association.

Stith, S. M., Green, N. M., Smith, D. B., & Ward, D. B. (2008). Marital satisfaction and marital discord as risk markers for intimate partner violence: A meta-analytic review. *Journal of family violence*, *23*(3), 149-160.

Thoresen, S., & Hjemdal, O. K. (2014). Vold og voldtekt i Norge Vol. 1. *En nasjonal forekomststudie av vold i et livsløpsperspektiv.*

Tilden, T., Håland, Å. T., Hunnes, K., Fossli, G., & Oanes, C. J. (2015). Utprøving av systematisk tilbakemelding i par- og familieterapi: Barrierer og utfordringer. *Fokus på familien, 43*, 292-312.

Tilden, T., Håland, Å. T., Hunnes, K., Lærum, K., Finset, A., & Gude, T. (2010). Internett-basert tilbakemelding i par- og familieterapi ; erfaringer med empiriinformert terapi. *Tidsskrift for Norsk psykologforening*.

Tilden, T., Wampold, B., Ulvenes, P. G., Zahl-Olsen, R., Hoffart, A., Barstad, B., . . . Håland, Å. T. (In press). Feedback in Couple and Family Therapy: A Randomized Clinical Trial. *Family Process.*

Todahl, J., & Walters, E. (2011). Universal screening for intimate partner violence: a systematic review. *Journal of Marital and Family Therapy, 37*(3), 355-369.

Vivian, D., & Malone, J. (1997). Relationship factors and depressive symptomatology associated with mild and severe husband-to-wife physical aggression. *Violence and Victims, 12*(1), 3.

Walker, L. E. (1979). The battered woman. New York, NY: *Harper & Row*.

Weede, E. (1981). Income inequality, average income, and domestic violence. *Journal of Conflict Resolution*, 639-654.

Wolf, K. A., & Foshee, V. A. (2003). Family violence, anger expression styles, and adolescent dating violence. *Journal of Family Violence, 18*(6), 309-316.

Yeh, H.-C., Lorenz, F. O., Wickrama, K., Conger, R. D., & Elder Jr, G. H. (2006). Relationships among sexual satisfaction, marital quality, and marital instability at midlife. *Journal of Family Psychology, 20*(2), 339.

Young, M., Luquis, R., Denny, G., & Young, T. (1998). Correlates of sexual satisfaction in marriage. *The Canadian Journal of Human Sexuality, 7*(2), 115-128.

Zahl-Olsen, R., Gausel, N., Tilden, T., & Håland, Å. T. (In review). Monitoring therapeutic change through diversity or simplicity? A conservative, critical Confirmatory Factor Analysis test of a Routine Outcome Monitoring system.

Zahl-Olsen, R., & Oanes, C. J. (2017). An Anthill of Questions that Made Me Prepare for the First Session: A Clinical Vignette of the Usage of STIC Feedback System. In T. Tilden & B. Wampold (Eds.), *Routine Outcome Monitoring in Couple and Family Therapy* (pp. 189-209). Cham, Switzerland: Springer.

Zinbarg, R. E., Pinsof, W. M., Quirk, K., Kendall, A., Goldsmith, J., Hardy, N., . . . Latta, T. (2018). Testing the convergent and discriminant validity of the Systemic Therapy Inventory of Change Initial scales. *Psychotherapy Research, 28*, 734-749. doi:https://doi.org/10.1080/10503307.2017.1325022

Øverlien, C. (2012). *Vold i hjemmet - barns strategier*. Oslo: Universitetsforlaget.

Øverlien, C., & Holt, S. (2018). Letter to the Editor: Research on Children Experiencing Domestic Violence. *Journal of Family Violence*. doi:10.1007/s10896-018-9997-9

Table 1
List of items used to model Physical Violence

	Variable	Item name	Items in the STIC questionnaire
Physical Violence	Couple violence		We get into shoving or hitting each other when we fight
	Family violence		Someone in my family is physically abusive to other family members
			There is someone in my family who pushes other family members around physically to get his or her way
Predictors	Anger		I am filled with anger toward my partner
	Expectation	Expect 1	My partner often acts like he or she can't stand me
		Expect 2	My partner often complains that I don't do my share of work around the house
		Expect 3	I am expected to do too much.
	Sexual satisfaction	Sexsat 1	I am sexually frustrated in this relationship
		Sexsat 2	I am sexually satisfied with my partner
	Self-control	Selfcont 1	Had urges or impulses that you could not control
		Selfcont 2	Had fits of rage you could not control

Table 2
Scale inter-correlations and descriptive statistics

	1	2	3	4	5	6
1 Expectation	-					
2 Anger	.31**	-				
3 Sexual satisfaction	-.36**	-.37**	-			
4 Self-control	-.11**	-.31**	.09**	-		
5 Family Violence	.21**	.12**	.00	-.11**	-	
6 Couple violence	.30**	.25**	-.05	-.17**	.28**	-
Mean	3.41	3.68	2.71	1.69	4.70	4.74
SD	.82	.99	1.06	.74	.57	.54
α	.65	-	.81	.61	.85	-

*Note: **p< .01.*

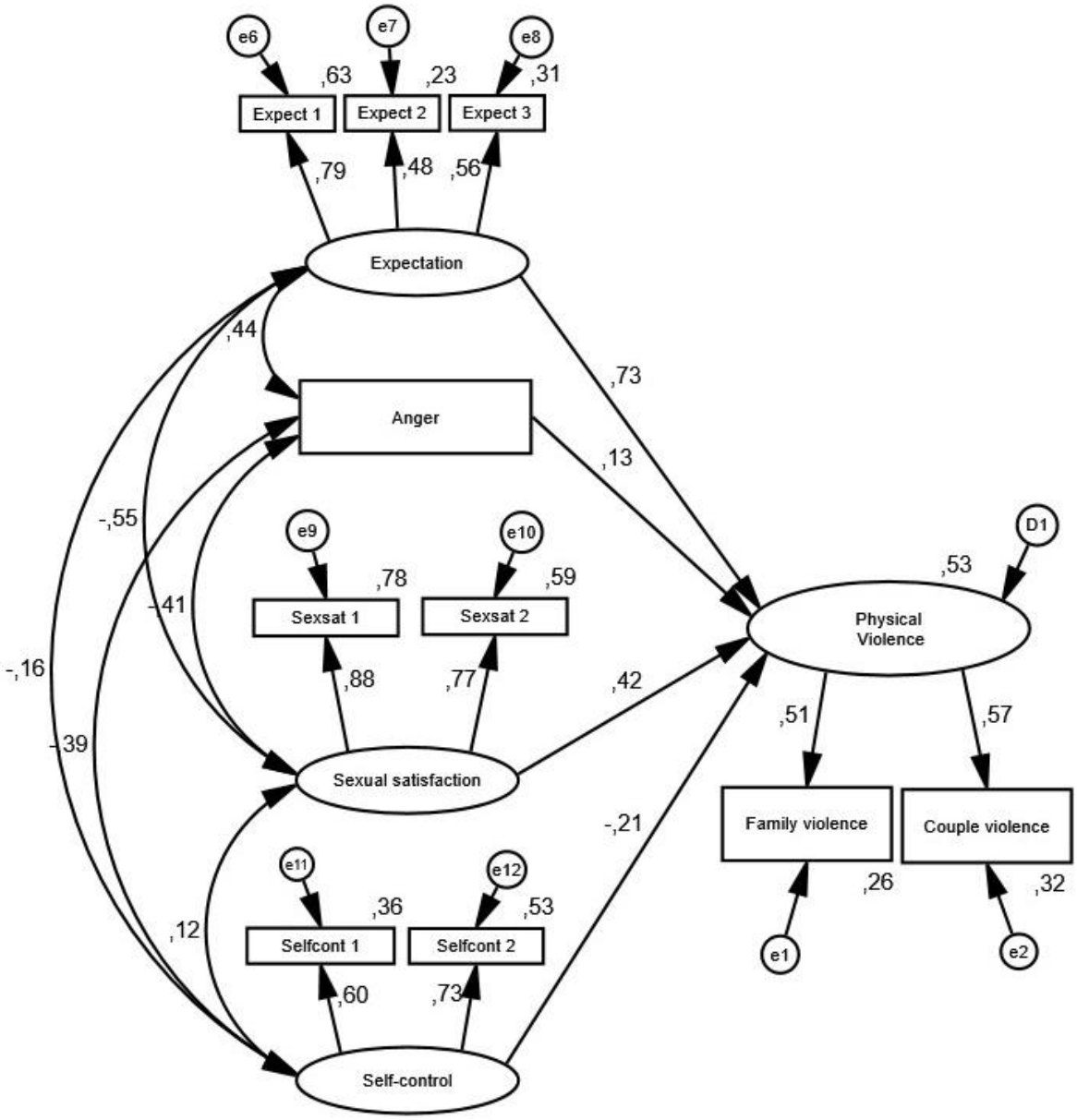
Supplementary table

Characteristics of the total sample and the differences between the sample with and without physical couple violence

	All		Physical couple violence		No physical couple violence				
	N = 830		N = 169		N = 604				
Variable	%	N	%	N	%	N	sig	χ^2	df
Sex: Female	51.8	430	47.3	80	52.5	317	.237	1.400	1
Education							.736	.612	2
Low	52.5	436	54.4	92	52.0	314			
Medium (Bachelor)	31.6	262	31.4	53	32.0	193			
High (Master or PhD)	15.2	126	13.6	23	15.9	96			
Income							.001	13.612	2
Low (400 000 NOK and below)	12.7	105	16.0	27	8.9	54			
Medium (401 000 to 800 000 NOK)	54.2	450	59.8	101	53.8	325			
High (above 800 000 NOK)	32.0	266	23.7	40	36.6	221			
Income sufficient	88.1	731	85.2	144	90.1	544	.073	3.212	1
Relationship status							.653	.202	1
Comitted relationship - not married	19.5	162	18.9	32	20.7	125			
Married	71.1	590	75.1	127	74.3	449			
Medication									
Using medication (one or more)	21.7	180	18.9	32	22.5	136	.318	.996	1
Depression	13.7	114	11.8	20	14.2	86	.104	2.640	1
Anxiety	7.0	58	9.5	16	6.3	38	.480	.500	1
Concentration difficulties/hyperactivity	2.5	21	1.2	2	2.8	17	.139	2.185	1
Bipolar	3.4	28	3.6	6	3.6	22	.631	.231	1
Other	6.0	50	8.3	14	5.1	31	.376	.784	1
Prior experience with therapy							.043	8.165	3
None	25.3	210	17.8	30	27.3	165			
Less than one year	37.1	308	43.8	74	36.8	222			
One to three years	20.5	170	18.3	31	20.4	123			
More than three years	16.4	136	19.5	33	15.4	93			
Family Violence Dicotomous	24.9	207	49.7	84	18.0	109	.000	84.324	1
Couple Violence Dicotomous	20.4	169	100.0	169	0.0	0			

Figure 1

Predictive model of Physical Violence



Paper 3

Zahl-Olsen, R., Håland, Å.T., Gausel, N., Wampold, B., Tilden, T. (2019). Change in work functioning from pre- to post-treatment in feedback-informed couple and family therapy in Norway. *Journal of Family Therapy*. https://doi.org/10.1111/1467-6427.12283

journal of family therapy

Journal of Family Therapy (2019) **0**: 1–2
doi: 10.1111/1467-6427.12283

Change in work functioning from pre- to post-treatment in feedback-informed Couple and Family Therapy in Norway

Rune Zahl-Olsen [a], Åshild Tellefsen Håland,[a] Nicolay Gausel,[b] Bruce Wampold[c] and Terje Tilden[c]

Work functioning has significance for the individual and society but has rarely been used as an outcome measure for psychotherapy. Work-related factors such as work satisfaction and working hours impact personal and relational life. More than half of those on sick leave suffer from social problems such as family-related distress or mental health issues rather than medical issues. This article investigates work functioning change from pre- to post-treatment in feedback-informed couple and family therapy. With a sample of 165 clients from different parts of Norway, we used hierarchical multiple regression and calculated clinical significant change in the analysis. We found firstly that work functioning improved from pre- to post-treatment. Secondly, we found that level of depressive symptoms, couple distress and family functioning predicted work functioning at pre-treatment. Thirdly, we found that the improvements on these measures (depressive symptoms, couple distress and family functioning) predicted work functioning at post-treatment.

Practitioner points

- Clients in couple and family therapy improve from start to end of therapy in work functioning, depressive symptoms, couple distress and family functioning
- Individual and relational improvement predicts the level of work functioning at post-treatment
- Therapists should attend to the client's difficulties not only at individual and relational levels but also in life functioning
- Even if effect sizes are high for the treatment, not all clients improve

Keywords: couple and family therapy; improvement; prediction; reliable change; work functioning

[a] Department of child and adolescent mental health, Sorlandet Hospital, Kristiansand, Norway
[b] University College Østfold, Fredrikstad, Norway
[c] Modum Bad Research Institute, Vikersund, Norway

Introduction

Dropping out of work can have severe consequences for the individual and society. It is established knowledge that occupational functioning is associated with mental health and relationships with one's family. For instance, the individual consequences of dropping out of work may be reduced personal wellbeing, including poorer finances (Warr, 2003), depression (Adler *et al.*, 2006; Lerner and Henke, 2008), relationship distress (Novak *et al.*, 2017) and reduced family functioning (Rotunda *et al.*, 1995). Dropping out of work not only has consequences for individual and family life, but also work-related factors such as low work satisfaction, long hours and a mobile office have been found to have a negative impact on couple relationships (Tavistock Centre for Family Relationships, 2018). Reciprocally, the experience of personal and relational distress has been found to reduce a person's work functioning (Whisman and Uebelacker, 2006).

For society, the welfare costs associated with those dropping out of work are great in most countries of the Organization for Economic Co-operation and Development (OECD) (Setzer and Rürup, 2013). Recent research performed in Scandinavia found that more than 50 per cent of sick leave was due to social problems (e.g. family-related distress or mental health issues) rather than medical issues (Aronsson *et al.*, 2015). The proportion of those who go on sick leave due to mental health issues has been steadily increasing in recent years (Hensing *et al.*, 2006) throughout the OECD countries (OECD, 2003). For instance, psychiatric disorders are now the most common reason for long-term sick leave (Henderson *et al.*, 2011). For this reason, psychotherapy should be one of several means in the effort to reduce the interrelated problems involving work, mental health, and relational distress.

People suffering from depression, anxiety (Bilsker *et al.*, 2005) and relational discord (Aronsson *et al.*, 2015) also have an elevated risk of impaired work functioning. In a special issue of the journal *Family Process* focusing on empirically supported treatments in couple and family therapy (CFT), the editor concludes: 'these treatments now extend to treat a wide array of significant couple and family problems, suggesting the value of couple and family therapy for both relational problems and problems often thought of as nested within individuals' (Lebow, 2016, p. 387). More specifically, CFT has been found to be effective for treating depression and anxiety (Carr, 2014a; Crane and Payne, 2011; Gurman, 2008, 2015), as well as dyadic and family problems (Sexton *et al.*, 2013). Hence one may assume CFT to be a suitable method in the

effort to increase clients' levels of functioning at work by reducing their depressive and anxiety symptoms and improving their relationships. To our knowledge, this is the first study that explicitly examines work functioning in the field of CFT. Therefore, we see work functioning as an area of interest to be more thoroughly examined through research. The study presented in this article contributes to addressing this objective.

This article raises the following research questions:

(1) To what extent is there a change in work functioning for clients from pre- to post- CFT treatment?
(2) To what extent is there a change in depressive symptoms, dyadic adjustment and family functioning from pre- to post- CFT treatment?
(3) To what extent is clients' work functioning at pre-treatment associated with depressive symptoms, dyadic adjustment and family functioning at the same time point?
(4) To what extent is clients' work functioning at post-treatment associated with change in depressive symptoms, dyadic adjustment, and family functioning from pre- to post- CFT treatment?

Method

Participating units

This study is a sub-study based on a naturalistic randomised clinical trial (RCT) (the 'main study', $n = 229$: Tilden *et al.*, in review), conducted at three levels of CFT sites in Norway. The main study's aim was to explore whether the use of the feedback system Systemic Therapy Inventory of Change (STIC; Pinsof *et al.*, 2009) made a difference compared to treatment as usual. In this current sub-study, not all data from the main study was included. However, this study included three CFT sites, each representing one of the main CFT service levels in Norway. Our rationale for using multiple sites was to capture the variation of how CFT is practised in our country, hence strengthening the ecological validity of the study, which would facilitate the generalisability of the findings. In line with a stepped-care model, the Step 1 CFT site in this study is a low-threshold family counselling outpatient service located in two Norwegian cities of medium size, where no referral is needed. Step 2 is represented by a mid-threshold inpatient/outpatient child and adolescent psychiatry unit in a general hospital in a medium-sized city in Norway, where a referral is needed. The third step is

represented by a high-threshold residential family treatment unit in a highly specialised national psychiatric hospital in Norway, where again a referral is needed.

Due to the great variation of client characteristics and problems, the treatments were tailored to the single client/couple/family's needs, often from an integrative theoretical approach (Pinsof, 1983). All the therapists were trained in using the STIC (Pinsof *et al.*, 2009). However, the therapists did not know the focus of this study.

Sample

As shown in Table 1, a total of 165 adult clients were included in this study. The data collection started in August 2013 and ended in September 2016. The mean age of the clients was 40.3 (± 7.8), and 52.1 per cent were women.

Measurements

The Short Form Quality of Life measure (SF-36 v.2) is a health-related quality of life client self-assessment measure consisting of thirty-six items. For the purposes of this article, we only make use of the data from the Role Emotional scale (SF-RE) that measures the client's work functioning. For the analyses of change, we needed a cut-off between normal and impaired work functioning. No such cut-off level was established for the work functioning measure (SF-RE). Thus, we applied the Jacobson and Truax (1991) formula by comparing the means and standard deviations of our sample with a Norwegian normative sample (Loge and Kaasa, 1998): cut-off = $(s_0M_1 + s_1M_0)/(s_0 + s_1)$. Here, M_0 and s_0 are the mean and standard deviation of the normative sample, and M_1 and s_1 are the mean and standard deviation of our sample, resulting in a cut-off value for work functioning at 79.35. A client's score below this cut-off value could clinically be helpful when hypothesising about the client's emotional problems as associated with less time spent on work, accomplishing less at work, or that the work is performed less carefully than usual. Answering 'most of the time' on at least one of these items or another combination that leads to the same combined average is required to obtain a score below the cut-off. Such a level of work functioning implies that people's emotional problems impact impaired work functioning in one or all these three ways. In our sample, Cronbach's alphas for the SF-RE were .91 (95 per cent confidence interval CI: .88 – .93) at pre-treatment and .92 (CI: .88 – .94) at post-treatment.

TABLE 1 *Characteristics of the sample*

	n = 165	
Variables	%	*n*
Sex: female	52.1	86
Education		
Low	49.1	81
Medium (Bachelor's degree)	33.3	55
High (Master's and PhD)	17.6	29
Relationship status		
Committed relationship – not married	20.6	34
Married	68.5	113
Medication		
Using medication (some more than one medication)	17.6	29
Depression	10.3	17
Anxiety	4.8	8
Concentration difficulties/hyperactivity	3.6	6
Bipolar	4.2	7
Other	4.8	8
Prior experience with therapy		
None	32.7	54
Less than a year	33.3	55
More than one year	33.9	56

The Beck Depression Inventory (BDI-II – (Beck *et al.*, 1996)) is a 21-item client self-report instrument for measuring the severity of depression, with a clinical cut-off at 14.29. Cronbach's alpha at pre-treatment was .93 (CI: .91 – .95) and .95 (CI: .94 – 96) at post-treatment. A clinical level of depression is defined as BDI> 14.29.

The Family Assessment Device (FAD; Epstein *et al.*, 1983) general functioning scale is a client self-assessment measure of perceived family functioning with twelve items, with a clinical cut-off at 2.00. Cronbach's alpha at pre-treatment was .89 (CI: .86 – .91) and at post-treatment it was .91 (CI: .88 – .94).

The Revised Dyadic Adjustment Scale (RDAS; Busby *et al.*, 1995) is a widely used 14-item client self-assessment questionnaire that provides a global measure of an individual's assessed level of consensus, satisfaction, and

cohesion with his or her spouse. The clinical cut-off is 47. Cronbach's alpha at pre-treatment was .85 (CI: .82 – .89) and at post-treatment .88 (CI: .82 – .91).

Procedures

Clients who agreed to participate in this study completed the Outcome Package (OP), meaning filling out the above-presented questionnaires as well as the initial STIC questionnaire electronically via secure zone internet before and after treatment. Additionally, the clients completed the intersession STIC before each therapy session. All data were received, analysed and securely stored by the Psychotherapychange.org website at www.family-institute.org, and then a report of this information was available to the therapists on the therapist website immediately after completion. This was presented in a graphical format and with summaries so that the information was easy for the therapists to use (Zahl-Olsen and Oanes, 2017).

Clients who did not complete the OP immediately upon the termination of treatment received three emails or phone call reminders before they were marked as missing.

Analyses

We used IBM SPSS v. 24 for the statistical analysis. None of the outcome measures violated the assumptions for parametric analysis after application of the central limit theorem to assume normal sample distribution (Field, 2018). Paired t-tests and effect sizes were calculated to explore statistical change between pre- (t1) and post- (t2) treatment. We calculated effect sizes both by dividing the difference of the means by the mean standard deviation of the two measures (Cohen's *d*) and by dividing the differences of the means by the standard deviation at post- (t2) treatment (Glass Δ). Analyses of correlations were performed to explore hypothesised associations between variables. Confidence intervals were calculated using 1000 bootstrapped samples. Based on systemic theory (Johnsen and Torsteinsson, 2012), assuming that personal, dyadic and family aspects mutually affect each other and considering the findings in the correlation analysis, we carried out a total of five hierarchical multiple regression analyses with forced entry with work functioning as the dependent variable.

In the first regression, we analysed whether the individual and relational measures at pre-treatment (t1) would predict the level of work functioning at the same time point (t1). In the second, third and fourth

regression we predicted work functioning at end of treatment (t2) by each of the other measures individually, firstly at pre-treatment (t1) after controlling for level of work functioning at pre-treatment (t1). Secondly, we analysed whether change on each measure from pre- to post-treatment predicted work functioning at post-treatment (t2) when controlling for work functioning at pre-treatment (t1). In the final and fifth regression analysis, we did the same as in regression two, three and four but this time including all measures. First, we analysed if the levels of all the measures together at pre-treatment (t1) predicted work functioning at post-treatment (t2) after controlling for level of work functioning at pre-treatment (t1). Secondly, we analysed whether the change on these measures from pre- to post-treatment predicted work functioning at post-treatment (t2) when controlling for work functioning at pre-treatment (t1). However, due to power analysis estimates these analyses were limited to being conducted on the total sample only, a result also supported by the rule of thumb $n = 50 + 8m$ where m is the number of variables in the regression analysis (Tabachnick and Fidell, 2014).

We also wished to assess the outcome on an individual level by use of the clinical significance (CS) approach (Jacobson and Truax, 1991), categorising clients into four outcome groups: recovered, improved, unchanged and deteriorated, on work functioning. Only those who at pre-treatment (t1) had impaired work functioning due to emotional problems were included in the CS calculation. We calculated the reliable change index (RCI) by using the formula provided by Jacobson and Truax (1991), as follows: RCI = 1.96 * SD $\sqrt{2(1-r)}$. Here, SD is the standard deviation, and r is the Cronbach's alpha. As Jacobson and Truax (1991) did not specify which sample to take the SD and Cronbach's alpha from, we chose to use those from our impaired sample in the RCI equation and found RCI to be 20.47 for the work functioning. To be categorised in the recovered outcome group it required a change from pre- (t1) to post-treatment (t2) that crossed the measurement's cut-off level (79.35) between impaired work functioning and normal work functioning and that this change was larger than the RCI.

Clients that completed the study

Of the 165 clients in the study, ninety-six completed the OP at termination. Of the sixty-nine that did not complete the termination OP, nineteen were outpatients still involved in treatment at the time of closing the data collection and were therefore not asked to complete the

termination OP. The omission of these clients yields a completion rate of 65.8 per cent (96/146).[1]

We performed attrition analysis by running one-way ANOVA analyses for the scale measures and Chi-square tests for the nominal measures. In general, on seventeen variables we found no significant difference between those who completed the final OP and those who did not. However, there were significant differences on the following two variables: significantly more of those who were married χ^2 (1, $n = 165$) = 7.864, $p = .005$, $phi = -.218$ and those who had prior experience with therapy χ^2 (1, $n = 165$) = 8.221, $p = .016$, $phi = .223$ completed the OP at termination.

Ethics

Informed consent for collecting the project data was obtained from each participant. This study was approved by the Modum Bad Ombudsman for Data Protection and the Regional Ethics Committee for Medical Research with human subjects (REK – no. 2017/96). The main study is registered at ClinicalTrials.gov.

Results

The research questions addressed outcomes on work functioning (SF-RE), depressive symptoms (BDI), family functioning (FAD), and couple distress (RDAS).

Change and effect size

As shown in Table 2, we found a statistically significant improvement from the start to the end of treatment for work functioning ($d = 0.34$), depression ($d = 0.38$), family functioning ($d = 0.48$) and couple distress ($d = 0.43$). For comparison with other studies, we additionally calculated pre- to post-effect size for those clients who were in the clinical/impaired region for each of the measures at pre-treatment (t1).

Reliable change

By use of the clinical significance (CS) approach (Jacobson and Truax, 1991) on work functioning, we found that the portion of clients who had impaired work functioning (<79.35) at pre-treatment (t1) represented

[1]Little's test of the hypothesis that data are missing completely at random (MCAR) is non-significant. This suggests that treating the data as MCAR may be a reasonable assumption and will not have a material impact on the accuracy of the estimates and inferences.

TABLE 2 *Levels of distress and outcomes*

			t1		t2					
Sample	Measure	N	M	SD	M	SD	t	Δ	*d*	Sig.
Total sample	Work functioning	97	73.63	24.61	81.79	23.36	–3.261	0.34	0.34	<.001
	Depression	92	15.1	11.38	10.78	11.28	4.160	0.38	0.38	<.001
	Couple distress	82	39.51	10.57	43.72	8.81	–5.149	0.48	0.43	<.001
	Family functioning	89	2.35	0.56	2.08	0.56	5.116	0.48	0.48	<.001
Clinical/ impaired sample	Work functioning	54	56.17	19.10	72.53	26.68	–4.100	0.61	0.71	<.001
	Depression	45	24.84	7.63	16.27	12.84	4.915	0.67	0.84	<.001
	Couple distress	66	35.73	7.52	41.17	7.56	–5.916	0.72	0.72	<.001
	Family functioning	68	2.59	0.39	2.22	0.54	5.958	0.69	0.80	<.001

Note: Higher values indicate less distress for Work function and Couple distress while it is the opposite for the other measures.

52.7 per cent ($n = 87$) of the total sample. As shown in Table 3, 53.7 per cent of the clients improved or even recovered at a reliable level from start to end of treatment.

Correlations

We conducted two correlation analyses; one on the data at the start of therapy and one on the data at the end of therapy (see Table 4). Significant associations were identified between work functioning and depression, work functioning and family functioning, but not for work functioning and couple distress at start or end of treatment.

Regression analyses

We wanted to investigate how to predict work functioning by addressing different levels of a client's life based on the individual, couple and family measures used in this study. In the first regression, we assessed if the prediction of work functioning improved when we added relational measures to the individual measure (pre-treatment prediction). Secondly, and more importantly for this study, we wanted to predict work functioning at post-treatment by the change from pre- to post-treatment of the individual and relational measures (post-treatment prediction). Statistical tests indicated that multi-collinearity was not a significant problem. Variance inflation factors (VIFs) were computed for each predictor variable. As a rule of thumb, a VIF> 10 indicates problematic collinearity (Field, 2018). The VIF in our data was below 3.2 and thus well below the suggested cut-off value.

Pre-treatment prediction. We used hierarchical multiple regression to assess the level of three measures (depression, couple distress and family functioning) to predict the levels of work functioning at pre-treatment (t1). Table 5 shows the results.

The level of depression was entered at Step 1, explaining 46.8 per cent of the variance in the work functioning. Couple distress was added

TABLE 3 *Reliable change*

Group	% (***n***)
Recovered	42.6 (23)
Improved	11.1 (6)
Unchanged	35.2 (19)
Deteriorated	11.1 (6)

TABLE 4 *Correlations*

Time	Measure	Work functioning	Depression	Couple distress	Family functioning
t1	Work functioning	–			
	Depression	–.684**	–		
	n	161			
	Couple distress	.077	–.319**	–	
	n	149	148		
	Family functioning	–.232**	.341**	–.621**	–
	n	152	151	142	
t2	Work functioning	–			
	Depression	–.646**	–		
	n	93			
	Couple distress	.165	–.447**	–	
	n	84	81		
	Family functioning	–.322**	.368**	–.717**	–
	n	91	87	78	

*Note: ** Correlation is significant at the 0.01 level (2-tailed). Higher values indicate less distress for Work function and couple distress while it is the opposite for the other measures.*

TABLE 5 *Hierarchical regression 1. Dependent variable: Work functioning (t1)*

	β	95 % CI	R Square Change	F Change	Sig.
Step 1			0.468	140.062	<.001
Depression (t1)	–1.608	–1.859 – –1.375			.001
Step 2			0.018	5.131	.025
Depression (t1)	–1.648	–1.922 – –1.417			.001
Couple distress (t1)	–0.389	–.750 – .031			.034
Step 3			0.019	5.589	.020
Depression (t1)	–1.688	–1.976 – –1.431			.001
Couple distress (t1)	–0.701	–1.065 – –0.295			.001
Family functioning (t1)	–8.113	–14.669 – –1.046			.020

Note: Confidence intervals are based on 1000 bootstrap samples.

at Step 2 and family functioning at Step 3. Both steps gained significant F change. The final model explains 53.4 per cent of the variance in work functioning at pre-treatment.

Post-treatment predictions. As shown in Table 6, we tested whether each of the measures (depression, couple distress, family functioning) by themselves could predict work functioning at post-treatment (t2) when work functioning at pre-treatment (t1) was controlled for. This was done using hierarchical multiple regression for each of the measures one by one. None of the measures at pre-treatment (t1) could significantly predict work functioning at post-treatment (t2). At the next step, we entered the level at post-treatment (t2) for each of the measures. Since we controlled for pre-treatment levels in Step 2, this gave us a residual change score for each of these measures. By themselves, each of the change scores could predict the level of work functioning at post-treatment (t2) when the level of work functioning at pre-treatment (t1) was controlled for.

Finally, we ran a hierarchical multiple regression in which we used all the measures together in one regression. In the first step, we controlled for work functioning at pre-treatment (Step 1). At Step 2 we added all the measures at pre-treatment (t1). As shown in Table 7 this model (Step 2) could not significantly predict the level of work functioning at end of treatment. At Step 3 all the measures at post-treatment (t2) were added, giving residual change score, showing that the model predicts 53.8 per cent of the level of work functioning at end of treatment.

Discussion

There are three main findings in this study. Firstly, we found improvements from pre- to post-treatment on work functioning. Secondly, we found that level of depression, couple distress and family functioning predict work functioning at pre-treatment. Thirdly, we found that the improvements on these measures (depression, couple distress and family functioning) predict work functioning at post-treatment.

To our knowledge, no previous study within CFT has addressed significant improvements from pre- to post-treatment on work functioning. When inspecting the group of clients with impaired work functioning at pre-treatment we found improvements on work functioning from pre- to post-treatment with close to large effect size. By use of the clinical significance (CS) approach (Jacobson and Truax, 1991), we found

TABLE 6 *Hierarchical regression 2, 3, 4. Work functioning at post-treatment (t2) predicted by each of the measures one by one*

	Depression					Couple distress					Family functioning				
	β	95 % CI	R Square Change	F Change	Sig.	β	95 % CI	R Square Change	F Change	Sig.	β	95 % CI	R Square Change	F Change	Sig.
Step 1			.251	30.179	<.001			.240	25.216	<.001			.217	24.175	<.001
Work func. (t1)	.480	.290 –.716			.001	.469	.246 – .716			.001	.473	.246 – .744			.001
Step 2			.007	.783	.379			.000	.000	.991			.009	.987	.323
Work func. (t1)	.407	.005 – .124			.147	.469	.236 –.724			.002	.444	.214 – .707			.002
a (t1)	–.225	.434 – –.778			.273	–.002	–.429 –.485			.990	-4.156	–14.462 – 5.452			.390
Step 3			.242	42.580	<.001			.052	5.771	.019			.058	6.950	.010
Work func. (t1)	.376	.166 – .626			.002	.490	.269 – .712			.001	.450	.238 – .689			.002
a (t1)	.510	.069 –.943			.033	–.512	–1.104 – .179			.119	3.655	–9.434 – 14.665			.551
a (t2)	–1.281	–1.824 – –.741			.001	.833	.004 – 1.483			.029	–12.768	–22.687 – –1.921			.016

Note: In the left pane a = Depression, in the middle pane a = Couple distress, in the right pane a = Family functioning. Confidence intervals are based on 1000 bootstrap samples.

TABLE 7 *Hierarchical regression 5. Work functioning at post-treatment (t2) predicted by all other measures combined*

		Change statistics				
Model	R Square	R Square Change	F Change	df1	df2	Sig. F Change
Step 1	.203	.203	18.063	1	71	.001
Step 2	.217	.014	.405	3	68	.750
Step 3	.538	.321	15.064	3	65	.001

Note: Step 1: Predictors: (Constant), work functioning (t1). Step 2: Predictors: (Constant), work functioning (t1), Depression (t1), Couple distress (t1), Family functioning (t1). Step 3: Predictors: (Constant), work functioning (t1), Depression (t1), Couple distress (t1), Family functioning (t1), Depression (t2), Couple distress (t2), Family functioning (t2).

that 53.7 per cent of this group had significant improvements on work functioning when measured at post-treatment, a finding similar to what is typical for other measures of improvement in psychotherapy (Ogles, 2013). We also found that 46.3 per cent were unchanged or even deteriorated from pre- to post-treatment on work functioning. This finding is important in at least two ways. Firstly, it should work as a reminder to clinicians that not all individuals are helped in treatment even if there is a significant change from beginning to end for the total sample of clients. This implies an ethical imperative for therapists to make a stronger effort to look for signs of no change or deterioration during treatment to optimise the therapy outcome. For researchers, this finding implies a methodological objective, addressing the need to measure outcome on an individual as well as on a group level. Secondly, it suggests that we need more knowledge about the group of clients who do not benefit from treatment. Even though there is a growing field of knowledge on non-responders (see, e.g., Day *et al.*, 2014; Mohr, 1995) to our knowledge this has not so far been investigated within the field of CFT.

This study also found that depressive symptoms predict the level of work functioning at pre-treatment. We also found that when we expanded the model to add couple distress and finally family functioning the prediction became more precise for each step. This finding supports the systemic theory claiming that different areas of life impact each other and a broader context is needed to understand a phenomenon (Johnsen and Torsteinsson, 2012). The most powerful predictor of work functioning at pre-treatment was the level of depression. This is in line with previous research implying that the level of depressive symptoms is associated with work performance (Adler *et al.*, 2006; Kessler *et*

al., 2006; Lerner and Henke, 2008; Mintz *et al.*, 1992). Research showing this relationship the other way around is sparse. However, Aronsson *et al.* (2015) and Dahl, Hansen and Vignes (2015) give some empirical support that couple distress and family functioning impact work functioning. In our study, we found that depressive symptoms and couple distress were interrelated, hence we suggest that these variables may interact mutually. Further research should explore in greater detail which mediators and change mechanisms influence the relationship between CFT and improved work function. The clinical implication of this prediction of work functioning at pre-treatment is that the therapist prior to or at the start of therapy should assess the client's difficulties on the individual, relational and functional levels to obtain optimal understanding as a basis for choosing and discussing interventions with clients.

As described, we also investigated whether the change on individual and relational levels predicted work functioning at post-treatment. From pre- to post-treatment, clients improved on all the included measures. More importantly for this study, the improvement in individual and relational aspects of clients' lives predicts the level of work functioning at post-treatment separately (depressive symptoms, couple distress, family functioning) as well as combined. We cannot from this study conclude that therapy alone affected improved work functioning, even if such an interpretation has support from the Tavistock Centre for Family Relationships report (2018). It is still reasonable to assume that therapy plays a role in this improvement. An even stronger finding is that these variables seem to be related, something that supports the essence of systemic theory; that individual and relational issues are related and a change in one variable is considered to create change among the interrelated variables (Johnsen and Torsteinsson, 2012). The clinical implication is, therefore, that which variable to focus on first in therapy may not be decisive. However, our results show that the individual level of depression answers to some of the variance in work functioning that is different from the variance explained by couple distress and family functioning. Hence, contrary to systemic theory, this implies that clinically it may matter which variable to give priority to at a certain point in the therapy. We find this interpretation in line with what Pinsof and Lebow (2005) labelled 'differential causality', namely that even though the systemic rule of thumb is to consider interactions as mutually influencing each other, different variables in such an interaction represent variance in strength, distance and impact. This perspective also yields support from network theory (Borsboom, 2017) that builds on creating idiographic network

maps based on data from the individual client. Thus, variables in this map (for instance, symptoms, behaviour, relationships) may relate to each other with different closeness, strength and centrality. In this way the therapist and client are given a therapeutic tool pointing to which variable(s) to focus on in order to optimise the desired change in the system. Hence, considering elements within relationships as unidirectional should not be perceived as contrary to a systemic approach.

This study also identified significant improvement in depressive symptoms, couple distress and family functioning. Improvement in depressive symptoms is in line with the findings of Carr (2014a) who identified several CFT treatments that are effective for mental health issues, including depression. This is different from the earlier review by Sexton and colleagues (2003), who identified a limited impact of CFT on depression. The other two improvements, couple distress and family functioning, are as expected since these objectives represent the explicit foci within CFT accompanied by empirical support (Carr, 2014b; Sexton, Datchi *et al.*, 2013; Sexton, Robbins *et al.*, 2003). The effect sizes we calculated on the total sample are small to medium, something that is similar to what has been found in other CFT studies (Sexton *et al.*, 2013). A comment regarding these effect sizes: the total sample includes a great variety of clients reporting from minor to major distress. This combination may very well appear within one couple or family, as it is well known to CFT therapists that one spouse might report more distress than the other. Methodologically, this variation is levelled out by including the total sample when calculating effect size, masking the severity of the most distressed clients. To illustrate, when selecting a subsample of the most impaired clients in our study – something that would be comparable to individual psychotherapy – we found effect sizes in the medium to high range, similar to psychotherapy in general (Ogles, 2013). This phenomenon is important to emphasise, as effect sizes are the standard reference for comparing effectiveness in psychotherapy. Hence, if calculating effect sizes without taking the mentioned variation within CFT into consideration, these effect sizes may not be comparable to individual psychotherapy.

Strengths, limitations and future research

Including data from all three levels of CFT care in Norway as well as using a naturalistic design strengthens the ecological validity of our study. Unfortunately, our sample size did not allow for analyses of

possible differences between the sites. In this study, we did not implement treatment manuals; hence, we did not control for adherence and competence. Even though this is a strength from a naturalistic standpoint, it is also a limitation because the variability of the treatment (such as therapeutic focus, specific methods, quality of alliance) may be large. However, we did control for the therapists' theoretical identification and nearly all of the therapists identified themselves as practising family/systems therapy. In retrospect, we regret not having video recordings of the therapies, which could have shed light on these objectives. The effect sizes in our study, showing outcomes like those of other studies, suggest that the delivered therapy, in general, was of satisfactory quality. All the included cases had the component of feedback given through the STIC system. It is a limitation of our work discussed here that we did not control for how the therapists used the feedback throughout the course of therapy. Although the effect sizes were good, RC analysis showed that many did not benefit from the treatment. We therefore suggest further research on the group of clients who do not benefit from CFT treatment, especially since classifying clients as responders and non-responders can be problematic (Senn, 2018). We also suggest further research to explore in detail the relationship between CFT and work functioning (i.e. mediators and change mechanisms).

Acknowledgement

This project was funded by Sorlandet Kompetansefond and Sparebanken Sor, Norway.

References

Adler, D. A., McLaughlin, T. J., Rogers, W. H., Chang, H., *et al.* (2006) Job performance deficits due to depression. *American Journal of Psychiatry*, **163**(9): 1569–1576.

Aronsson, G., Johansen, V., Marklund, S., Rønning, R. *et al.* (2015) *Sjukfrånvarons dimensioner: svensk-norska jämförelser och analyser.* Stockholm: Liber.

Beck, A. T., Steer, R. A. and Brown, G. K. (1996) Beck depression inventory-II. *San Antonio*, **78**(2): 490–498.

Bilsker, D., Gilbert, M., Myette, T. L. and Stewart-Patterson, C. (2005) *Depression & work function: bridging the gap between mental health care & the workplace.* Vancouver: Mental Health Evaluation & Community Consultation Unit at the University of British Columbia. Retrieved from https://www.sfu.ca/carmha/publications/depression-and-work-function.html

Borsboom, D. (2017) A network theory of mental disorders. *World Psychiatry*, **16**(1): 5–13. https://doi.org/10.1002/wps.20375

Busby, D. M., Christensen, C., Crane, D. R. and Larson, J. H. (1995) A revision of the dyadic adjustment scale for use with distressed and nondistressed couples: construct hierarchy and multidimensional scales. *Journal of Marital and Family Therapy*, **21**(3): 289–308. https://doi.org/10.1111/j.1752-0606.1995.tb00163.x

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

Carr, A. (2014b) The evidence base for family therapy and systemic interventions for child-focused problems. *Journal of Family Therapy*, **36**(2): 107–157. https://doi.org/10.1111/1467-6427.12032

Crane, D. R. and Payne, S. H. (2011) Individual versus family psychotherapy in managed care: comparing the costs of treatment by the mental health professions. *Journal of Marital and Family Therapy*, **37**(3): 273–289.

Dahl, S. Å., Hansen, H. T. and Vignes, B. (2015) His, her, or their divorce? Marital dissolution and sickness absence in Norway. *Journal of Marriage and Family*, **77**(2): 461–479.

Day, M. A., Thorn, B. E. and Rubin, N. J. (2014) Mindfulness-based cognitive therapy for the treatment of headache pain: a mixed-methods analysis comparing treatment responders and treatment non-responders. *Complementary Therapies in Medicine*, **22**(2): 278–285.

Epstein, N. B., Baldwin, L. M. and Bishop, D. S. (1983) The McMaster family assessment device. *Journal of Marital and Family Therapy*, **9**(2): 171–180.

Field, A. (2018) *Discovering statistics using SPS* (5th ed.). London: Sage Publications.

Gurman, A. S. (2008) A framework for the comparative study of couple therapy. In A. S. Gurman (ed.) *Clinical handbook of couple therapy* (pp. 1–26). New York: Guilford Press.

Gurman, A. S. (2015) The theory and practice of couple therapy. In A. S. Gurman (ed.) *Clinical handbook of couple therapy* (5th ed.) (pp. 1–18). New York: Guilford Press.

Henderson, M., Harvey, S. B., Øverland, S., Mykletun, A., *et al.* (2011) Work and common psychiatric disorders. *Journal of the Royal Society of Medicine*, **104**(5): 198–207. https://doi.org/10.1258/jrsm.2011.100231

Hensing, G., Andersson, L. and Brage, S. (2006) Increase in sickness absence with psychiatric diagnosis in Norway: a general population-based epidemiologic study of age, gender and regional distribution. *BMC Medicine*, **4**(1): 19.

Jacobson, N. S. and Truax, P. (1991) Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, **59**(1): 12–19.

Johnsen, A. and Torsteinsson, V. W. (2012) *Lærebok i familieterapi*. Oslo: Universitetsforl.

Kessler, R. C., Akiskal, H. S., Ames, M., Birnbaum, H., *et al.* (2006) Prevalence and effects of mood disorders on work performance in a nationally representative sample of US workers. *American Journal of Psychiatry*, **163**(9): 1561–1568.

Lebow, J. L. (2016) Editorial: Empirically supported treatments in couple and family therapy. *Family Process*, **55**(3): 385–389. https://doi.org/10.1111/famp.12240

Lerner, D. and Henke, R. M. (2008) What does research tell us about depression, job performance, and work productivity? *Journal of Occupational and Environmental Medicine*, **50**(4): 401–410. https://doi.org/10.1097/JOM.0b013e31816bae50

Loge, J. H. and Kaasa, S. (1998) Short form 36 (SF-36) health survey: normative data from the general Norwegian population. *Scandinavian Journal of Social Medicine*, **26**(4): 250–258.

Mintz, J., Mintz, L., Arruda, M. and Hwang, S. (1992) Treatments of depression and the functional capacity to work. *Archives of General Psychiatry*, **49**(10): 761–768. https://doi.org/10.1001/archpsyc.1992.01820100005001

Mohr, D. C. (1995) Negative outcome in psychotherapy: a critical review. *Clinical Psychology: Science and Practice*, **2**(1): 1–27.

Novak, J. R., Sandberg, J. G. and Davis, S. Y. (2017) The role of attachment behaviors in the link between relationship satisfaction and depression in clinical couples: implications for clinical practice. *Journal of Marital and Family Therapy*, **43**(2): 352–363. https://doi.org/10.1111/jmft.12201

OECD (2003) *Transforming disability into ability. Policies to promote work and income security for disabled people*. Paris: OECD Publications Service.

Ogles, B. M. (2013) Measuring change in psychotherapy research. In M. J. Lambert (ed.) *Bergin and Garfield's handbook of psychotherapy research and behavior change* (6th ed.) (pp. 134–166). New Jersey: Wiley.

Pinsof, W. M. (1983) Integrative problem-centered therapy: toward the synthesis of family and individual psychotherapies. *Journal of Marital and Family Therapy*, **9**(1): 19–35.

Pinsof, W. M. and Lebow, J. L. (2005) A scientific paradigm for family psychology. In W. M. Pinsofand J. L. Lebow (eds.) *The art of the science* (pp. 3–19). New York: Oxford University Press.

Pinsof, W. M., Zinbarg, R. E., Lebow, J., Knobloch-Fedders, L. M., *et al.* (2009) Laying the foundation for progress research in family, couple, and individual therapy: the development and psychometric features of the initial Systemic Therapy Inventory of Change. *Psychotherapy Research*, **19**(2): 143–156.

Rotunda, R. J., Scherer, D. G., Imm, P. S. and Deleon, P. H. (1995) Family systems and alcohol misuse: research on the effects of alcoholism on family functioning and effective family interventions. *Professional Psychology: Research and Practice*, **26**(1): 95–104. https://doi.org/10.1037/0735-7028.26.1.95

Senn, S. (2018) Statistical pitfalls of personalized medicine. *Nature*, **563**(7733): 619–621. https://doi.org/10.1038/d41586-018-07535-2

Setzer, M. and Rürup, B. (2013) Globalization and welfare — Consequences for unemployment, income distribution and welfare. In U. Steger (ed.) *Facetten der Globalisierung: Ökonomische, soziale und politische Aspekte* (pp. 119–132). Berlin: Springer-Verlag.

Sexton, T. L., Datchi, C., Evans, L., LaFollette, J., *et al.* (2013) The effectiveness of couple and family-based clinical interventions. In M. J. Lambert (ed.) *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed.) (pp. 587–639). New Jersey: John Wiley & Sons.

Sexton, T. L., Robbins, M. S., Hollimon, A. S., Mease, A. L., *et al.* (2003) Efficacy, effectiveness, and change mechanisms in couple and family therapy. *Handbook of family therapy: the science and practice of working with families and couples* (pp. 229–261). New York, NY: Brunner-Routledge.

Tabachnick, B. G. and Fidell, L. S. (2014) *Using multivariate statistics* (6th ed.). Harlow: Pearson Education Limited.

Tavistock Centre for Family Relationships (2018) *Couple relationships and work; work and couple relationships: a policy briefing from the Relationships Alliance*. Tavistock Relationships Alliance: Retrieved from https://tavistockrelationships.ac.uk/images/uploads/Couple_Relationships_and_Work_RA_briefing.pdf

Tilden, T., Wampold, B., Ulvenes, P. G., Zahl-Olsen, R., *et al.* (2019). Feedback in couple and family therapy: a randomized clinical trial. *Family Process*.

Warr, P. (2003) Well-being and the workplace. *Well-being: Foundations of Hedonic Psychology*, p. 392.

Whisman, M. A. and Uebelacker, L. A. (2006) Impairment and distress associated with relationship discord in a national sample of married of cohabitating adults. *Journal of Family Psychology*, **20**(3): 369–371.

Zahl-Olsen, R. and Oanes, C. J. (2017) An anthill of questions that made me prepare for the first session: a clinical vignette of the usage of STIC feedback system. In T. Tilden and B. Wampold (eds.) *Routine outcome monitoring in couple and family therapy* (pp. 189–209). Cham, Switzerland: Springer.