Mapping Individual Symptom Networks for Post-Traumatic Stress Disorder

Elizabeth Blomfield

January 2022

Research submitted in partial fulfilment of the requirements for the degree of Doctor in Clinical Psychology (DClinPsy), Royal Holloway, University of London.

Acknowledgments

I would like to thank Dr. Gary Brown and Dr. Sharif El-Leithy for their guidance, support, patience and kindness over the extended period during which I completed this thesis. I am also grateful to all the participants, both clinicians and clients, who so generously contributed their time and experiences to this project.

The debt I owe to my parents is immense and inexpressible. I will always be grateful to them for their love, humour and constant encouragement. My mother Clare was my earliest inspiration in psychology, as an Educational Psychologist who clearly loved her work and excelled at it. My father Stephen fired me with a love of words and writing, and encouraged me to do what brought me joy. I hope they would both be glad and proud to see me complete this project.

My wonderful sisters, Louise, Camilla and Antonia have always supported me and helped me not to take myself too seriously. I am lucky to have been sustained by many incredible friends at times when it felt as though I would never finish, especially Alex, Jenny, Piers, Julie, Anna, Eldyne, and Shalini. My profound thanks to Anais and Antonia for all the childcare, without which I would still be writing this thesis for several more years. Finally, the biggest thanks, hugs and cheers to Alex, Theo and Phoebe. My children have barely known a time when I wasn't working on this thesis. They won't know what has hit them! Alex, thank you for your unfailing support and love throughout this long period of writing and research.

This thesis is in memory of my beloved Mum and Dad, and for Alex, Theo and Phoebe, with love and gratitude.

Table of Contents

List of Tables and Figures

List of Tables

Paper One: Systematic Review

Paper Two: Empirical Study

List of Figures

Paper One: Systematic Review

Paper Two: Empirical Study

Lay Summary

Systematic Review

One of the key tools used by clinical psychologists working with people in psychological distress is clinical case formulation. Formulation is the name given to the explanation a clinical psychologist develops about why a person has come to have psychological difficulties, what is keeping those difficulties going and how it may be possible to relieve them. Despite its importance, and the fact that clinical psychologists try to ensure their work is based on evidence, there is surprisingly little research into the use of formulation, including whether a formulation is effective and contributes to recovery. This may be because no consensus has been reached on appropriate criteria for evaluating formulation.

Reliability and validity are two criteria which have been used to assess formulations. Reliability in case formulation means that two different clinicians would produce the same formulation for the same client. Validity is commonly described as how well a measure does what it claims it is doing, so a valid case formulation would describe accurately what is contributing to a person's difficulties. The tools clinical psychologists use in research have to be reliable and valid, so it seems reasonable to assess whether clinical tools, such as formulation, also meet that standard. The first part of this thesis, the systematic review, aimed to evaluate what research has shown so far about whether cognitive and behavioural case formulation is reliable and valid. We also considered whether assessing formulation quality may be another useful approach to formulation evaluation. To carry out this review, we searched research databases for papers which examined the reliability and validity of cognitive and/or behavioural formulation. Thirty papers were selected as relevant and reviewed to determine what they said about reliability, validity and quality. Our conclusions were:

1. <u>Reliability</u>
    i. Definitions of reliability varied substantially.

ii. Different formulations were found to agree, but often only for categories such as describing symptoms, rather than for underlying symptom causes, or possible treatment.

2. Validity
   i. Definitions of validity also varied substantially.
   ii. Studies often measured validity indirectly, using aspects such whether the client's symptoms improved to indicate whether the formulation was valid.
3. Quality
   i. Quality assessment appeared to be a more objective and comprehensive evaluation method than either reliability or validity.
   ii. We were unable to review the quality of formulations within the study as insufficient raw data was provided.
4. Conclusions
   i. Reliability and validity were not particularly useful ways of evaluating formulation. Even if clinicians produced different formulations for the same person, both might assist recovery. Validity was important but studies often used methods which evaluated validity indirectly or at the end of treatment, which was too late to be useful clinically, although assessments of validity for use within treatment were developing.
   ii. Quality assessment was a promising approach to formulation evaluation, but more research into how formulation quality affected outcomes was required in order to confirm the best criteria to use.

Empirical Study

Network analysis is a relatively new approach to understanding mental health difficulties. Instead of conceptualising mental health difficulties as having an underlying cause such as depression, network analysis suggests that depression as a mental state is produced by interactions between symptoms such as feeling sad and lacking motivation.

This study looked at two types of networks of symptoms in a small sample of people with post-traumatic stress disorder (PTSD). The networks we looked at were the network of symptoms found to occur roughly at the same time, and the network of symptoms that seemed to follow on from one another. We wanted to answer two questions:

1) If there was a relationship over time between cognitive symptoms and emotional symptoms. According to one theoretical idea of how PTSD develops, known as the cognitive theory of PTSD, this relationship maintains PTSD symptoms and we wanted to evaluate if these relationships could be identified using network analysis.
2) Whether network analysis and a formulation created by a clinician for the same person's mental health difficulties resembled each other in terms of complexity and relationships between the symptoms.

We found:

1) It was not possible to confirm whether there was a relationship between cognitive and emotional symptoms within PTSD as too few pairs of these symptoms were found to draw firm conclusions.
2) Network analysis and clinicians' formulation overlapped in several ways, but overall, clinicians' formulations were more complex and provided better explanations about how difficulties developed and were maintained. All the relationships identified in network analysis-derived formulations were specific to that person, whereas this was less clear for clinicians' formulations.

While our conclusions were tentative, the study showed that network analysis was a useful approach to answering these questions and that by personalising the variables included, further studies could provide more insight into the relationship between cognitive and emotional symptoms. Our study also suggested that network analysis might be useful in clinical settings, as a

complement to clinical formulation which could provide more detail about specific relationships between symptoms.

Integration, Impact and Dissemination

Both the systematic review and empirical study considered what the qualities of a good formulation may be and how this could be achieved. The systematic review focused on understanding which criteria might be most appropriate for evaluating formulation while the empirical study compared formulations developed through network analysis and by clinicians in order to evaluate their relative merits and limitations in terms of providing insight into an individual's difficulties. The findings of the systematic review suggested that the most helpful criteria, such as those included in quality assessment, encompassed both what the formulation included and how well it developed an overall explanation of difficulties. The empirical paper proposed that formulations from both clinicians and network analysis could provide useful information to help with treating psychological difficulties. While the process of developing the thesis had not been smooth, the experience had highlighted issues such as the need to balance the demands of network analysis on clients with its likely contribution to a better recovery from psychological distress.

A Systematic Review of the Reliability and Validity of Formulation in Cognitive and Behavioural Therapy

Abstract

Formulation may be defined as a tool through which knowledge of psychological theory and research is applied to the specific mental health difficulties of an individual in order to determine appropriate intervention. Its importance, both as a key skill for clinical psychologists, and evidence of the empiricism on which their practice is based, has been acknowledged by professional bodies and leading figures within academic psychology. It is surprising therefore that the reliability and validity of such a foundational practice within clinical psychology remains unclear. This review aimed to consider the ways in which validity and reliability have been defined within cognitive and behavioural therapy, the degree to which it has been found reliable and valid according to those definitions, and the limitations of those assessments in providing an adequate evaluation of the value and role of formulation.

A systematic literature search was conducted in January 2021 for English-language studies evaluating the reliability and/or validity of formulation for adults within cognitive or behavioural therapy. Thirty articles met inclusion criteria.

The review found that achieving good levels of inter-rater reliability between formulations was difficult, often found only in discrete contexts, and dependent to a notable degree on agreement with specific definitions of reliability, which undermined the use of reliability as a neutral psychometric term. Operationalising validity in a way that permitted direct assessment was also challenging, with proxies for validity often substituted, limiting insight. Overall, conclusions were that use of the psychometric measures reliability and validity was problematic in the context of formulation, although the need to demonstrate that formulation is underpinned by scientific rigour was acknowledged. Assessment of formulation quality was proposed as a direction for future research into the means by which objective standards for formulation and so a stronger evidence base for its use may be established.

Formulation is an essential skill for clinical psychologists, not only cited as one of the five core skills of psychology which underpins professional practice (British Psychological Society [BPS], 2017) but heralded as the "lynchpin that holds theory and practice together" (Butler, 1998, p.1), the "therapist's compass" (Persons, 1989, p. 37), which is "critical" to understanding client difficulties (Beck et al., 1979, p.78) and "the heart of evidence-based practice' (Kuyken et al., 2005, p. 1188). As a profession, clinical psychologists proclaim themselves "more than psychological therapists; they are scientist practitioners" (Division of Clinical Psychology [DCP], 2010, p.2), and a significant amount of psychological practice takes place within a National Health Service where evidence-based practice is considered fundamental (The Information Standard, 2013). Given these facts, it is surprising, therefore, that there is perceived to be a paucity of empirical research into formulation, and that evidence for its contribution to clinical outcomes remains unclear (Bieling & Kuyken, 2003; Flinn et al., 2014).

Definitions of Formulation

The variety of practice within psychology has led to a range of approaches to and definitions of formulation. However, there is consensus around the idea of formulation as a tool through which psychologists apply their broad knowledge of psychological theory to the specific difficulties experienced by a particular person, in order to determine an appropriate and theoretically justified intervention:

> "Psychological formulation is the summation and integration of the knowledge that is acquired by this assessment process that may involve psychological, biological and systemic factors and procedures. The formulation will draw on psychological theory and research to provide a framework for describing a client's problem or needs, how it developed and is being maintained…. Psychological intervention, if considered appropriate, is based on the formulation." (DCP, 2010, pp.5-6).

The flexibility of formulation, which allows room for interpretation and clinical expertise alongside the application of psychological theory, means it can be especially useful in work with complex

presentations or when standardised treatments are ineffective (Davidson, 2008). An important aspect of psychological formulations is that they are hypotheses. As such, they are tentative, and open to revision by therapist and client as therapy proceeds (Leahy & Beck, 1996; Persons, 1989). Indeed, this "process of ongoing collaborative sense-making" (Harper & Moss, 2003, p.8) between therapist and client is a key feature of formulation in many approaches, especially cognitive therapy.

Formulations are also strongly reflective of the theory from which they are derived. Psychologists draw on multiple theoretical perspectives in their practice, and so multiple models for formulation co-exist, drawing on different types of information, or using different explanations to justify their recommended intervention. A psychodynamic approach may look for the origins of difficulties in recurring relational patterns based on unconscious processes, while a systemic formulation may seek explanations for current difficulties in the client's socio-cultural context. In addition, formulation may be focused at the problem level, on a specific issue or situation, or at the case level, which attempts to develop a more comprehensive explanation for all of a client's difficulties (Persons & Tompkins, 1997).

Difficulties in establishing the scientific basis for formulation may be tied to its status as a skill which marries objective empirical evidence and more subjective clinical judgement, although subjectivity within scientific research must also be acknowledged. Historically, the emergence of cognitive case formulation in particular is tied to the development of the 'scientist-practitioner' model within clinical psychology, a term originally used as a description of the balance of research and clinical practice endorsed as a training program for clinical psychologists by the American Psychological Association [APA] in 1924, and supported at the Boulder Conference in 1949 (Kennedy & Llewellyn, 2001). The importance to the profession of maintaining scientific credibility was made clear by Touyz (1995) who argued that to lose its scientific practitioner mantle would leave clinical psychologists indistinguishable from counsellors or psychotherapists.

The Reliability and Validity of Cognitive Formulation

If clinical psychologists wish to advocate for themselves as scientist-practitioners, and claim formulation as an evidence-based approach, then fundamental questions over its reliability and validity must be resolved. Being both reliable and valid is considered a prerequisite for basic psychometric measures used in cognitive research or therapy, yet the reliability and validity of formulation remains obscure. Without this quality being known, it is difficult to develop an understanding of how formulation contributes to good outcomes, how to ensure it is consistently effective and how to build on this understanding to improve its effectiveness.

Reliability can be defined as the quality of producing consistent results (Graziano & Raulin, 2014) so that, for example, a reliable formulation approach might be expected to lead to similar case conceptualisations by two different therapists assessing the same person (inter-rater reliability). Alternatively, a reliable formulation approach would also be expected to produce a similar case conceptualisation for the same person presenting at different times (test-retest reliability), provided their circumstances and difficulties were unchanged. Validity is more complex, but represents an assessment of the degree of "methodological and/or conceptual soundness of the research" (Graziano & Raulin, 2014, p.119), commonly summarised as a gauge of the how well a measure measures what it claims to measure. The validity of a measure is often evaluated by quantifying how well it predicts another variable. Neither reliability nor validity are considered absolute qualities; rather degrees of being reliable or valid are determined.

It is perhaps not surprising that these questions have been explored within the context of cognitive and cognitive behavioural formulation. Cognitive theory, on which cognitive behavioural therapy (CBT) and cognitive therapy (CT) are grounded, has accumulated a strong evidence base over time. Multiple studies have found support for different aspects of the cognitive theory of depression (Clark & Beck, 1999; Kwon & Oei, 1994; Teasdale, 1983). A similarly strong case has been built for the cognitive theory of anxiety (Rapee, 1991; Roth & Heimberg, 2001), although the evidence for a cognitive theory of personality disorder is more mixed (Barnow et al., 2009; Kramer et

al., 2013). Convincing data for efficacy supports the use of cognitive theory-based therapies such as CBT with anxiety, depression and obsessive-compulsive disorder (National Institute for Health and Care Excellence, [NICE] 2005, 2009, 2011) and cognitive therapy has become one of the most widely taught therapies, with training in CBT a core requirement of doctoral training courses in clinical psychology in the UK (Health & Care Professions Council [HCPC], 2015). The lack of evidence for the reliability and validity of cognitive and cognitive behavioural case formulation therefore represents a significant gap in otherwise persuasive support for CT and CBT.

Research into Reliability and Validity within Cognitive Therapy

Defining validity and reliability in relation to cognitive and cognitive behavioural case formulation is not straightforward, and research to date has been marked by differences of approach to the characterization and evaluation of these central concepts. These differences appear connected to use of different models of cognitive and cognitive behavioural case formulation, which focus on different aspects of a person's presentation or different levels of formulation. Models used in research include those based on the work of Aaron Beck and Judith Beck, for example the J. Beck Case Conceptualisation Diagram (CCD; Beck, 1995). Using this model, clinicians interrogate a client's developmental history and discuss with them their responses to a number of exemplar situation-thought-emotion-behaviour situations in order to infer deeper core beliefs, dysfunctional assumptions or rules, and consequential maladaptive compensatory strategies which are likely to maintain their difficulties. Another influential model of formulation is Persons' case formulation approach (1989). This describes psychological difficulties in terms of two levels; an upper level of overt difficulties (comprised of cognitions, behaviours and moods, which are interdependent) and an underlying hypothesized psychological mechanism, defined as a problem or deficit which produces the overt difficulties. Muran and Segal's (1992) model of self-schemas differs from the preceding models in focusing on situation-level, rather than case-level, formulation.

Reliability

In terms of reliability, what has been operationalised almost without exception in the research literature is degree of inter-rater agreement with an 'expert' formulation or benchmark. Given the variety of 'expert' models examined, this has led to competing claims for reliability. Within this context, common themes have emerged, although few definitive answers.

Multiple research studies (Dudley et al., 2010; Persons et al., 1995) and reviews (Bieling & Kuyken, 2003) have highlighted findings that inter-rater reliability for cognitive formulation can be good to substantial for the descriptive aspects of formulation, but notably weaker for the inferential elements. Persons and colleagues (1995) tested the hypothesis that clinicians could agree on both the overt problems and underlying cognitive mechanisms for two depressed and anxious clients using the Persons (1989) case formulation model. Participant clinicians listened to an audiotape of an initial assessment interview. Identification of most of the overt problems listed by 'expert judges' was high (>80%). Inter-rater reliability for identification of schemas and dysfunctional attitudes was poor when single judges were compared, although moderate when pooled across five judges (single judge intraclass correlation coefficient (ICC) range = 0.07-0.55 , five judge ICC excluding dysfunctional attitudes range = 0.66-0.92) . Eells and colleagues (1998) found similarly high agreement among kappa ratings for descriptive aspects of the formulation (*m* = 0.88), with inferred content either missing (mean inference rating for formulation was 1.80 (SD = 0.77) on a scale of 1 (descriptive) to 5 (highly inferential) or less reliable when included (e.g. strengths/adaptive skills κ = 0.73).

Factors including the provision of training in case conceptualisation, and additional structure/context for the formulation, have been proposed as factors likely to increase the reliability of formulation. However individual studies and recent systematic reviews (Easden & Kasantzis, 2018; Flinn et al., 2014) have produced only inconclusive support for this. In one study (Persons & Bertagnolli, 1999), training on the method of case conceptualisation was provided as well as a list of potential problem domains, anchor points and schema definitions. The inter-rater reliability of the

problem list was reasonable (mean percentage of problems agreed from 'expert list' = 67.46%) and ratings of schemas was good over five judges (ICC *m* = 0.72) though remained poor for single judges (ICC *m* = 0.37). However, analysis of variance showed no effect of the provision of specific context on schema identification. A further study (Mumma & Smith, 2001) provided heuristic guidelines for formulation within the context of a more challenging task which asked clinicians to generate prototypical examples of how clients might respond to different situations, referred to as cognitive-behavioural interpersonal scenarios (CBISs). This produced good inter-rater reliability for fourteen out of fifteen dimensions, although again, only once ratings were aggregated over all ten raters (ICC *m* = 0.83), with the clinicians acknowledging, although unable to prove directly, that the additional structure provided may have contributed to their positive results. Easden and Kazantzis (2018) noted that use of a clear rating system may have contributed to reliability in this study, but that it may also have been due to the greater contextual information provided by using audio recordings of clients rather than commonly used written vignettes. Kuyken and colleagues (2005) found evidence of tentative connection between training and/or experience and higher reliability in a study comparing the reliability and quality of case formulations produced by expert, experienced and novice clinicians, where pre-qualified clinicians' formulations were less reliable for more inferential information.

Validity

Assessments of validity within cognitive case formulation have taken a number of different routes towards evaluating how well formulation has succeeded in capturing, explaining and indicating ways to address a person's difficulties. Several studies have compared outcomes for formulation-driven versus manualised treatment. Some of these have found support for the idea that formulation-driven CBT produced at least comparable outcomes to standardised treatment. Persons et al. (2006) found significant reductions of symptoms of depression, with a large effect size, for conceptualization-driven CBT in private practice. These were in line with the outcomes listed in published randomised controlled trials with similar patients. Persons and Bertagnolli (1999)

conducted a comparison of conceptualisation-driven CBT versus CBT plus pharmacology for patients with depressed women and also found comparable outcomes for both groups. In Ghaderi (2006), a comparison of 'individualised' CBT incorporating logical functional analysis with standardised CBT for people with eating disorders showed improvement for both groups on the primary outcome measure, the Eating Disorders Examination (Cooper & Fairburn, 1987). Further analysis of objective bulimic episodes and compensatory behaviour following treatment showed superior results in the individualised condition, with a significant group difference in favour of the individualised condition for stopping excessive compensatory exercise. However, a recent systematic review (Easden & Kasantzis, 2018) noted design issues with studies to date which have used outcome as a measure of validity. First was the use of small samples, which limited the statistical power of studies to identify significant results, or raised questions of generalisability. Second, differences between formulation-driven and standardised treatments were often operationalised poorly, with therapists in the standardised treatments offering at least some degree of individualization, meaning that claims for the distinctive contribution of formulation were hard to justify.

An alternative approach to validity has been evaluation of its clinical utility for both clients and therapists. Principal findings from studies (Chadwick et al., 2003; Morberg-Pain et al., 2008) have been that therapists can find formulation useful, validating and that it increases their sense of alliance while it simultaneously produces positive, neutral and negative responses in clients, a far more mixed response. When Flitcroft and colleagues (2007) interviewed therapists, factor analysis of the responses revealed that quite different aspects of formulation were valued as most important by different therapists with the 'here and now' aspects of formulation, its role within the process of CBT and the 'trait', or longitudinal/developmental aspects all prized. While formulation is seen as useful, there seems to be low consensus about which aspects are most helpful, either between clients and therapists, or even between therapists. Without shared understanding about what formulations should contain or aim to do, agreement on their validity appears hard to achieve.

A New Systematic Review

Concluding their systematic review of studies evaluating the reliability and validity of cognitive case formulation, Easden and Kazantzis (2018) claimed that many studies' use of purely psychometric definitions of reliability and validity had limited their findings. By using evaluation methods which treated formulation as akin to a one-off assessment rather than as a process both informing and informed by therapy, they argued that studies failed to capture how formulation was used in practice and therefore how it may have contributed to change. They also offered strong critique of studies' use of ecologically invalid source material, such as brief written vignettes. These shortcomings highlight a central problem which confronts researchers attempting to evaluate cognitive case formulation using empirical methods: How best to evaluate what formulation does and how well it achieves its goals in objective, measurable terms which do not, in the attempt, distort the idiographic, evolving nature of formulation so much that the evaluation itself becomes invalid. If the larger purpose of reliability and validity assessment is to determine whether these are useful criteria for the objective evaluation of formulation, but attempts using these means have been unsatisfactory for reasons which relate at least in part to the nature of the criteria used, the question also arises of whether more suitable methods of formulation evaluation might exist.

One possibility is to include assessment of formulation quality in empirical evaluations of cognitive case formulation. Previous research has commented on a potential relationship between formulation quality and outcome (Nattrass et al., 2015) and considered relationships between formulation quality, validity and/or reliability (Kendjelic & Eells, 2007; Kuyken et al., 2005), suggesting these concepts are interlinked. Comparative evaluation of novice and expert clinicians' formulations has also shown that quality assessment can provide useful insight into formulations, identifying differences which indicate the ways in which formulations could be improved (Dudley et al., 2015; Haarhoff et al., 2011). Considering formulation quality within empirical assessment of formulation would address a criticism by Easden and Kazantzis (2018), that studies within their review compared studies not clearly equivalent, as they lacked assessment of clinicians' skills. Being

less tied to an established psychometric definition, assessment of formulation quality may also be defined in ways that permit objective, quantifiable assessment of formulation by observational methods that do not interfere with formulation process. Quality assessment may offer, therefore, a more ecologically valid approach to establishing the value and role of formulation, creating a clearer evidence base by which to support or critique its use.

On this basis, a new systematic review is proposed. This review is intended first to acknowledge the multiple ways in which reliability and validity have been defined in relation to individual case formulation, and evaluate at minimum the degree to which formulation has been found to be reliable and valid according to these definitions. It will then go further to consider the degree to which these definitions are sufficient. The review will include single case studies not included in the most recent systematic review by Easden and Kazantzis (2018), due to evidence that these include new developments in validity assessment not previously considered. This review intends to interrogate further the adequacy of reliability and validity as the basis for the empirical assessment of formulation by evaluating the quality of formulations in included studies, either by using their own quality assessment where possible, or else via a new quality assessment tool. It is hoped that insights from this evaluation may clarify questions about the relative value of different approaches to the evaluation of cognitive and/or behavioural case formulation.

The research questions for this review will be:

1. To what degree has individual case formulation been established as reliable in cognitive and behavioural therapy outcome research?
2. To what degree has individual case formulation been established as valid in cognitive and behavioural therapy outcome research?

Method

Literature Search

In January 2021 a broad systematic search strategy was employed to identify all relevant primary studies and previous systematic reviews of reliability and validity within cognitive, cognitive

behavioural and behavioural therapy research. Scoping searches had shown that studies used diverse descriptions of the key concepts 'formulation', 'cognitive and/or behavioural therapy', 'reliability' and 'validity', with relevant studies of reliability and validity particularly difficult to capture using search terms. As a result, our strategy was to employ a broadly-defined initial search for studies exploring formulation within cognitive and/or behavioural therapy and to identify those exploring reliability and validity at the title/abstract screening stage. All study designs within quantitative and qualitative research were included.

Firstly, databases PsycINFO and Web of Science was searched within all text and all sources using Boolean/Phrase search mode in PsycINFO, all fields in Basic search for all sources for Web of Science, from inception until the search date. No restrictions were made regarding year of publication, country where the research took place or that the study was a peer-reviewed published article as it was considered that grey literature such as doctoral theses may provide useful insights in this area. Our search terms for formulation in these databases were ("Psych*formulation" OR "case conceptualisation" OR "case conceptualization" OR "case formulation"). In terms of cognitive and/or behavioural therapy, the list of therapies included in our search were derived from the latest edition of a relevant practice handbook, *The Handbook of Cognitive-Behavioural Therapy* (Dobson & Dozios, 2019). The search terms used in these two databases was (CBT OR "cognitive behavioural therapy" OR "cognitive behavioral therapy" OR "cognitive behaviour therapy" OR "cognitive behavior therapy" OR "cognitive therapy" OR "CT" OR "cognitive psychotherapy" OR "emotion-centred problem-solving therapy" OR "emotion-centered problem-solving therapy" OR "emotion centred problem solving therapy" OR "emotion centered problem solving therapy" OR "rational emotive behaviour therapy" OR "rational emotive behavior therapy" OR "schema therapy" OR "mindfulness" OR "acceptance and commitment therapy" OR ACT OR "dialectical behaviour therapy" OR "dialectical behavior therapy" OR DBT OR "behavioural activation" OR "behavioral activation" OR "behaviour activation" OR "behavior activation" OR "behaviour therapy" OR "behavior therapy"). The concepts were linked using Boolean term AND.

Secondly, the database PubMed was searched in all fields from inception until the search date for the same search terms. Again, no restrictions were made regarding year of publication, country where the research took place, or peer review. To fit with PubMed protocols, the search terms used were, for formulation: ("Psychological formulation" OR "psychiatric formulation" OR "case conceptualisation" OR "case conceptualization" OR "case formulation"). For cognitive and/or behavioural therapy, search terms were as per the previous two databases. Following all database searches, results were filtered to limit results to English-language studies.

In addition to these searches, the database of UK clinical trials (clinicaltrials.gov) was searched for relevant studies, with none identified. Reference lists of studies identified as relevant were reviewed and eleven papers added to the abstract screening. Following duplicate removal, titles and abstracts of papers identified by all search methods were initially screened for eligibility, after which full-text papers were retrieved.

A second reviewer screened 20% of papers selected for full-text review. Agreement regarding selection of studies for full-text review was 80%. Disagreements were resolved by discussion and included studies finalised. The study protocol was registered with PROSPERO (https://www.crd.york.ac.uk/PROSPERO/) with registration number CRD42021235235.

Eligibility Criteria

Studies were included if they met the following criteria:

1. Cognitive and/or behavioural therapy research (as defined by Dobson & Dozios, 2019)
2. Considered the reliability and/or validity of case formulations included in the study. Reliability was understood as the consistency of the individual formulation approach when compared across the same client formulated by different clinicians. Evaluation of reliability included use of at least one measure of agreement such as inter-rater reliability, and test-retest reliability, ideally using objective statistical methods such as percentage agreement or calculation of Cohen's kappa. Validity was understood as the degree to which an individual

formulation approach captured the factors which contributed to and/or maintained the presenting problem in a way which led to the identification of appropriate intervention.

3. General adult sample (≥ 18 years of age)
4. Written in English.

Data Extraction

Data was extracted using an electronic spreadsheet tailored to the purpose of this review. The categories of data extracted from qualitative and quantitative studies were derived from my analysis of categories used in previous, similar systematic reviews, in particular Easden and Kasantzis (2018), and were:

Sample Size

The sample size used in each study was noted in order to allow evaluation of any possible effects of sample size on the data.

Participant Characteristics

Information was extracted regarding participant characteristics, including age (mean and range), gender, study population from which the sample was drawn and clinical diagnoses.

Key Findings

Key findings regarding reliability and validity were extracted from each paper.

Study Evaluation Targets

In order to identify the prevalence of common targets for evaluation in studies examining reliability/validity, information regarding whether each study assessed inter-clinician agreement, used an expert/benchmark formulation, evaluated the use of formulation on therapy outcome, and considered clinical utility for the client and/or therapist was extracted.

Model Integrity

As degree of fidelity to the model of formulation/treatment used was considered likely to affect results, information was extracted from studies regarding whether model integrity was checked and/or training in the model was prescribed.

Intervention and Comparator/Control Characteristics

To allow comparison between studies based on these factors, information was extracted regarding the model of formulation used, whether formulation was based on multiple sessions, intervention used, if any, therapist experience, level of training and profession.

Studies Evaluating Reliability – Use of Systematized Approach

To address the question of whether using a systematized/structured approach increased formulation reliability, as previous research had suggested (e.g. Mumma & Smith, 2001), information was extracted regarding whether studies used this approach.

Studies Evaluating Validity – Format, Intensity, and Duration of Intervention

Information was extracted regarding the format, intensity and duration of the intervention, if this was included in studies evaluating validity, to permit consideration of its influence on outcomes which may form part of validity evaluation.

Evaluation Characteristics – Quantitative Studies

To allow comparison between quantitative studies based on these factors, information was extracted regarding the methods used to assess reliability and/or validity, the criteria used in reliability/validity assessment, and whether the measures used to assess reliability/validity were themselves validated.

Evaluation Characteristics – Qualitative Studies

To allow comparison between qualitative studies based on these factors, information was extracted regarding the analytical approach used, themes derived which were relevant to the perception or reality of the reliability of the formulation, and themes derived which were relevant to the perception/reality of the validity of the formulation.

Risk of Bias

The score for quality assessment based on the Mixed Methods Appraisal Tool (MMAT; Pluye & Hong, 2014) was included in the data extraction. In addition, to allow consideration of the effects

of other forms of bias not included within this tool, information regarding study level risk of bias, outcome level risk of bias, and publication bias was extracted.

100% of the data extraction was reviewed by a second reviewer. Agreement regarding the accuracy of the data collection was 99.13%. Any differences were resolved by discussion and any changes necessary made to the final data extraction.

Evaluation of Formulation Quality

Prior to the literature search, criteria for assessment of the quality of formulation were drawn up between the researcher, her supervisor, and an external expert (listed as headings in Table 6). Scoping searches confirmed that it was not possible to make fully independent assessments of the quality of formulations used in studies as raw formulation data was rarely reported. Instead, we noted where evaluation of formulation quality within included studies indicated our quality assessment criteria had been met. We also noted studies where assessment of formulation quality was absent.

Evaluation of Methodological Quality

Methodological quality of studies included were assessed using the MMAT (Pluye & Hong, 2014). This tool was chosen as the review included both quantitative and qualitative studies and it permitted comparative scoring of heterogeneous study designs, allowing easier comparison of quality. The tool was assessed for reliability and efficiency at the pilot stage (Pace et al., 2012) and agreement between the raters was found to be moderate to perfect for the MMAT criteria, and substantial in relation to overall quality scores based on intra-class correlations (Shrout & Fleiss, 1979).

Twenty per cent of the quality assessment was reviewed by a second reviewer. Agreement regarding the quality assessment was 100%. The final quality assessment was taken into consideration during synthesis of findings, with any caveats regarding the findings of studies rated as low in quality reported.

Strategy for Synthesis of Findings

Given the heterogeneity of included studies, a narrative synthesis approach was selected as the most appropriate method to integrate study findings. Where similarity of methods allowed, such as with the four qualitative studies included, a combined analysis of all themes identified was carried out.

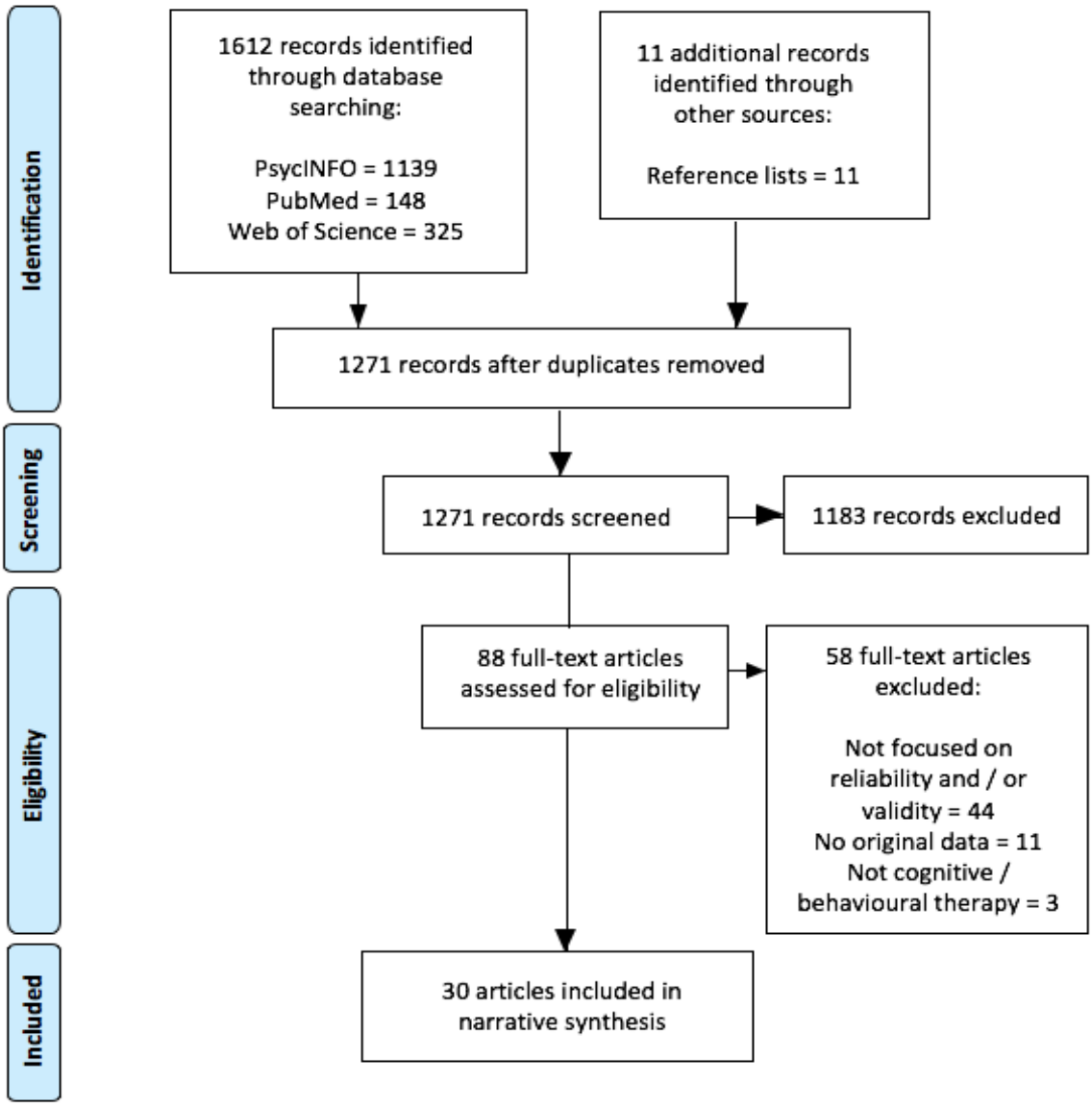
Results

Overview

Initial searches identified 1612 articles. Eleven additional articles were identified from reference lists (Bieling & Kuyken, 2003; Easden & Kasantzis, 2018; Redhead, Johnstone & Nightingale, 2015). Following duplicate removal, 1271 articles' titles and abstracts were screened for relevance and 1183 articles excluded. Eighty-eight full-text articles were retrieved, assessed for eligibility and 58 articles excluded. The study selection process and reasons for article exclusion at the stage of full-text review are documented in a PRISMA flow diagram (Figure 1). Thirty articles met eligibility criteria for this systematic review and were included in the narrative synthesis.

Figure 1

PRISMA Flow Diagram of Study Selection



Description of Included Studies

Details of included studies are stated in Table 1. For four categories of data included in the protocol, no relevant findings were reported so they were not included in results; the excluded categories were criteria used to assess formulation reliability, criteria used to assess formulation validity, therapist profession and whether formulation was based on multiple sessions. Study publication dates ranged from 1992 to 2018. All included studies were carried out in Western countries, with the majority conducted in the United States of America (*n* = 19, 63.3%), a further

eight conducted in the United Kingdom (26.7%), and one each in Sweden, New Zealand and Canada. Twenty-seven studies used a quantitative approach, including three randomised controlled trials, one non-randomised study (using an existing sample as a control), twenty-two quantitative descriptive studies and one study using mixed quantitative and qualitative methods. Three studies were qualitative evaluations of clients' experience of formulation.

Depending on the focus of the study, samples were either clinicians providing formulation or the clients for whom these formulations were provided. Quantitative study samples ranged from one to 60 clients (*m* = 19) or 23 to 115 clinicians (*m* = 55.89). Client samples were particularly small; eleven studies used samples smaller than ten (40.74% of quantitative studies) and six studies were demonstrations of an approach using a single client. Small sample size may have limited the conclusions studies were able to draw; in Ghaderi (2006) it was noted that a lack of statistical power reduced capacity to identify significant results, leading to possible Type II error. Clinician samples were larger but still small for the purpose of making between-group comparisons, for example between clinicians with differing levels of experience (as in Dudley et al., 2010). Qualitative study sample sizes were between seven and 13 clients.

Convenience sampling was dominant (*n* = 21, 70%) with purposive sampling used for studies requiring particular experience, for example advanced skills in case formulation (*n* = 9, 30%). The study population from which client samples were drawn were clinical outpatient samples for twelve studies, a clinical inpatient sample for one study and not stated for eight studies. No comparisons between samples and study populations were made, meaning the representativeness of samples was unclear.

Demographics of Included Studies

Client participants ranged in age from 18 to 70 (*m* = 35.08), with 63.53% of the sample female. Three studies did not report the age profile of client participants (Mumma & Fluck, 2016; Mumma et al., 2018; Nattrass et al., 2015). Three studies did not report the gender characteristics of client samples (Ghaderi, 2006; Hess, 2000; Kahlon et al., 2014). Included studies were strongly

focused on formulation for depression and/or anxiety, with this the primary clinical diagnosis considered in eighty percent of studies (*n* = 24). Other studies evaluated formulation for psychosis (*n* = 3), obsessive-compulsive disorder (*n* = 1), bulimia nervosa (*n* = 1) or did not state the diagnosis considered (*n* = 1).

Table 1

Characteristics of Studies and Demographics of Client Participants

		Sample size					Client participant characteristics			
Authors (date)	Country	Study design	Raters	Therapists	Client/ vignettes	Sampling method	Population	Age: mean (range)	Gender (% female)	Principal clinical diagnosis
Barris (1996)	USA	Randomised controlled trial	2	3 per patient	60	Convenience	Inpatient clinical	36.6 (18-70)	75	Depression
Chadwick, Williams & Mackenzie (2003)	UK	Mixed methods	1	2	Exp 1: 13 Exp 2: 4	Convenience	Outpatient clinical	Exp 1: 31.5 Exp 2: 39.3 (20-56)	Exp 1: 46 Exp 2: 50	Psychosis
Dudley, Park, James & Dodgson (2010)	UK	Quantitative descriptive		82	1 vignette	Purposive				Psychosis
Eells, Kendjelic & Lucas (1998)	USA	Quantitative descriptive	2	14	56	Convenience	Outpatient clinical	40 (20-66)	66.10	
Eells, Lombart, Kendjelic, Turner & Lucas (2005)	USA	Quantitative descriptive	6	65	6 vignettes	Purposive				GAD, MDD, BPD

		Sample size					Client participant characteristics			
Authors (date)	Country	Study design	Raters	Therapists	Client/ vignettes	Sampling method	Population	Age: mean (range)	Gender (% female)	Principal clinical diagnosis
Eells, Lombart, Salsman, Kendjelic, Schneiderman & Lucas (2011)	USA	Quantitative descriptive	6	65	6 vignettes	Purposive				GAD, MDD, BPD
Flitcroft, James, Freeston & Wood-Mitchell (2007)	UK	Quantitative descriptive		23	1 vignette	Convenience				Depression
Ghaderi (2006)	Sweden	Randomised controlled trial	2	1	48	Convenience	Outpatient clinical	27.2 (19-51)		Bulimia Nervosa
Haarhoff, Flett & Gibson (2011)	New Zealand	Quantitative descriptive		26	4 vignettes	Convenience				Depression / GAD
Hess (2000)	USA	Quantitative descriptive	4	3	7	Convenience	Outpatient clinical	24 (18-33)		Depression
Kahlon, Neal & Patterson (2014)	UK	Qualitative	2		7	Purposive	Outpatient clinical	33 (19-54)		Depression
Koerner (1997)	USA	Quantitative descriptive	6	4	17	Convenience	Outpatient clinical	41.1		MDD

		Sample size					Client participant characteristics			
Authors (date)	Country	Study design	Raters	Therapists	Client/ vignettes	Sampling method	Population	Age: mean (range)	Gender (% female)	Principal clinical diagnosis
Kuyken, Fothergill, Musa & Chadwick (2005)	UK	Quantitative descriptive	2	115	1 case study	Purposive				MDD and personality difficulties
Morberg Pain, Chadwick & Musa (2008)	UK	Qualitative	2	5	13	Purposive	Outpatient clinical	32.2 (21-52)	38	Psychosis
Mumma (2004)	USA	Quantitative descriptive	1	1	1	Convenience		44	100	MDD, Dysthymia
Mumma & Fluck (2016)	USA	Quantitative descriptive		1	1	Convenience		'Middle-aged'	0	GAD, MDD
Mumma & Mooney (2007)	USA	Quantitative descriptive		2	2	Convenience		43	0	MDD, Dysthymia, GAD
Mumma & Smith (2001)	USA	Quantitative descriptive	10	4	4	Convenience		47 (35-53)	100	MDD, GAD, PD
Mumma, Marshall & Mauer (2018)	USA	Quantitative descriptive		1	1	Convenience			0	Depression and anxiety

			Sample size					Client participant characteristics		
Authors (date)	Country	Study design	Raters	Therapists	Client/ vignettes	Sampling method	Population	Age: mean (range)	Gender (% female)	Principal clinical diagnosis
Muran, Samstag, Segal & Winston (1998)	USA	Quantitative descriptive	4	6	6	Convenience		41.83	67	Mood Disorder, Anxiety disorder, Personality Disorder
Muran, Samstag, Ventur, Segal & Winston (2001)	USA	Quantitative descriptive	2	1	1	Convenience		51	0	GAD, Avoidant Personality Disorder
Muran & Segal (1992)	USA	Quantitative descriptive	3	1	1	Convenience		41	0	Dysthymia, Anxiety Disorder
Muran, Segal & Samstag (1994)	Canada	Quantitative descriptive	2	5	8	Convenience	Outpatient clinical	39.3	60	Depression, Anxiety disorder, Social Phobia
Nattrass, Kellett, Hardy & Ricketts (2015)	UK	Quantitative descriptive	3	8	29	Convenience	Outpatient clinical		55	OCD

		Sample size					Client participant characteristics			
Authors (date)	Country	Study design	Raters	Therapists	Client/ vignettes	Sampling method	Population	Age: mean (range)	Gender (% female)	Principal clinical diagnosis
Persons & Bertagnolli (1999)	USA	Quantitative descriptive	2	47	3	Purposive	Outpatient clinical	37 (23-56)	100	Depression, GAD, Personality Disorder, Social Phobia
Persons, Bostrom & Bertagnolli (1999)	USA	Quantitative descriptive		1	45	Convenience	Outpatient clinical	32.1	60	MDD, BP, PD
Persons, Mooney & Padesky (1995)	USA	Quantitative descriptive	2	46	2	Purposive	Outpatient clinical		100	Depression and anxiety
Persons, Roberts, Zalecki & Brechwald (2006)	USA	Non-randomised quantitative study		4	58	Convenience	Outpatient clinical	36.4	60	Mood and/or anxiety disorder
Redhead, Johnstone & Nightingale (2015)	UK	Qualitative	2	7	10	Purposive	Outpatient clinical	(24-67)	80	Depression and/or anxiety

Note. Blank entry in table indicates relevant information was not provided by study; *MDD* Major Depressive Disorder; *GAD* Generalised Anxiety Disorder; *BPD* Borderline Personality Disorder; *PD* Panic Disorder; *OCD* Obsessive-Compulsive Disorder; *BP* Bipolar Disorder.

Evaluation of Reliability

Details of reliability assessment within included studies are listed in Table 2, with additional information in Table 3. Eighteen studies included evaluation of the reliability of cognitive case formulation (60%); of these, 14 considered both reliability and validity (46.67% of all included studies) and four reliability alone (13.33%).

In 16 studies, material for analysis was obtained by the study researchers asking clinicians to generate formulations in response to stimulus material. This material included live clinical interviews (*n* = 4), written vignettes (*n* = 4), videos of clinical interviews (*n* = 3), audiotapes of clinical interviews (*n* = 4) and session transcripts (*n* = 1). Two studies examined existing intake reports or clinic assessments. Most studies evaluated complete formulations (*n* = 12), although one study (Barris, 1996) focused solely on underlying cognitive mechanisms and five studies on self- or self- and interpersonal scenarios (Mumma & Smith, 2001; Muran & Segal, 1992; Muran et al., 1994; Muran et al., 1998; Muran et al., 2001).

Reliability was assessed almost exclusively in terms of inter-rater reliability (all 18 studies), with one study (Muran et al., 1998) also considering test-retest reliability. Expert views on 'correct' formulation were prominent: seven studies evaluated inter-rater reliability in terms of agreement with a benchmark formulation or problem list devised by experts, or with expert evaluation (Barris, 1996; Dudley et al., 2010; Haarhoff et al., 2011; Koerner, 1997; Kuyken et al., 2005; Persons et al., 1995; Persons & Bertagnolli, 1999). Five studies (Eells et al., 1998; Eells et al., 2005; Eells et al., 2011; Kendjelic & Eells, 2007; Nattrass et al., 2015) used agreement on assignment to the content categories of the Case Formulation Content Coding Manual [CFCCM] (Eells et al., 1998) as their measure of reliability. Five studies compared inter-rater agreement on the clinical relevance of self/interpersonal scenarios (Mumma & Smith, 2001; Muran & Segal, 1992; Muran et al., 1994; Muran et al., 1998; Muran et al., 2001) and one assessed agreement on allocation to condition, based on a case summary (Ghaderi, 2006).

Studies varied in the level of specificity they applied to assessment of formulation reliability, with some studies evaluating agreement at a fine level of detail and others limiting assessment to broader categories. The seven studies using expert formulations as benchmarks evaluated agreement precisely, assessing individual aspects of formulation such as core beliefs. Other studies considering reliability at a micro level included Mumma and Smith (2001), which assessed agreement between independently derived formulations on fifteen key dimensions, and the four studies by Muran and colleagues, which compared agreement between individual components of self-scenarios (Muran & Segal, 1992; Muran et al., 1994; Muran et al., 1998; Muran et al., 2001). The five studies which used the CFCCM exemplified a broader approach to reliability based on agreement at the level of categorisation of formulation content.

The most commonly used measure of agreement was the ICC (Shrout & Fleiss, 1979; eight studies), with Kappa coefficients used in six studies. These measures provided a more robust evaluation of agreement than percentage agreement, which fails to account for agreement by chance. Nonetheless, percentage agreement was used in seven included studies.

Key findings in relation to reliability assessment were, firstly, that it is possible to obtain good agreement for case formulation in terms of the broad categories of content included, as in Barris (1996), which found fair to good agreement for categorisation of a patient's underlying cognitive mechanism as sociotropic or autonomous (80% agreement, κ= 0.60). Studies using the CFCCM, for example Eells et al. (1998), also obtained good agreement across content categories (mean κ = 0.88, range = 0.71 - 1.0). However, reliability was reduced at a finer level of discrimination, with only poor agreement found for identification of specific underlying cognitive mechanisms (Barris, 1996). A similar pattern was found for inter-rater agreement, with multiple studies finding poor reliability between individual raters, but good reliability when averaged across groups (e.g. Koerner, 1997; Persons & Bertagnolli, 1999; Persons et al, 1995). Differences in the inter-rater reliability of different aspects of cognitive case formulation were also found, with descriptive and overt aspects such as behaviours and triggers consistently more reliably identified

than inferential elements such as dysfunctional assumptions and core beliefs (Dudley et al., 2010; Eells et al., 1998). Finally, studies found some evidence for factors increasing the reliability of formulation, for example higher qualifications (Persons & Bertagnolli, 1999) and experience or expertise (Dudley et al., 2010; Kuyken et al., 2005), although support for other factors also considered to increase reliability, such as increased structure and context during the formulation process, was less clear (Persons & Bertagnolli, 1999).

Table 2

Evaluation of Reliability

Authors (date)	Key findings	Use of systematized approach to formulation	Method of reliability assessment	Use of validated measure of reliability
Barris (1996)	• Fair to good inter-rater reliability of identification of patient's underlying cognitive mechanism as sociotropic or autonomous (80% agreement; κ = 0.60). • Poor inter-rater reliability of identification of specific underlying cognitive mechanism (20% agreement; κ = 0.05).	Yes – pre-determined categories of underlying cognitive mechanism	Percentage agreement and kappa coefficient	Yes
Dudley, Park, James & Dodgson (2010)	• Wide range in percentage agreement with items in benchmark CF (32.2-91.6%). • Highest reliability scores: overt presenting issues (e.g. behaviours 91.6%). • Lower agreement: inferential components (36.7-65.8% for core beliefs and dysfunctional assumptions). • Moderate agreement with experts' formulation among all raters (mean score 26.8/54, SD =5.88).	No	Percentage agreement with benchmark CF and score per item (0 = inaccurate to 2= accurate)	Yes
Eells, Lombart, Salsman, Kendjelic, Schneiderman & Lucas (2011)	• Reliability results same as Eells et al. (2005); same therapists, vignettes, transcriptions and content coding procedures used with different research questions.	No	Kappa coefficient for inter-rater reliability of content and quality ratings of CF	Yes

Authors (date)	Key findings	Use of systematized approach to formulation	Method of reliability assessment	Use of validated measure of reliability
Eells, Lombart, Kendjelic, Turner & Lucas (2005)	• Using CFCCM, multi-rater kappa coefficients for descriptive (κ = .61), diagnostic (κ = 0.81), inferential (κ = 0.62) and treatment (κ = 0.69) categories within CF obtained, reflecting good to excellent agreement beyond chance. • For case formulation quality, two-way random ICCs ranging from 0.69 - 0.89 found; rated good to excellent.	No	Kappa coefficient for inter-rater reliability of content categories and ICC for quality ratings of CF	Yes
Ghaderi (2006)	• 90% of detailed summaries of twenty randomly chosen cases assigned correctly to condition by three independent therapists.	No	Percentage agreement of correct assignment to one of two conditions	Yes
Haarhoff, Flett & Gibson (2011)	• Considered percentage agreement with expert formulation by novice clinicians as part of assessment of content and quality of formulation. • Average agreement across all four vignettes <30% for problem list. • Highest scores for percentage agreement with expert formulation: 'relevant childhood data', 'Axis I disorders', 'compensatory strategies' and 'core beliefs about self'. Very low agreement for 'therapy interfering behaviour' and 'Axis II personality disorder traits'. • Study assessed percentage of CFs which included items in different categories. Most frequently included item was 'inferred psychological mechanism' (94.5%). Low scores noted for 'inferred socio-cultural mechanisms' and 'therapy-interfering behaviours' (percentage not given).	No	Percentage agreement with expert formulation	Yes

Authors (date)	Key findings	Use of systematized approach to formulation	Method of reliability assessment	Use of validated measure of reliability
Kendjelic & Eells (2007)	• Randomised controlled trial comparing clinician formulations following CF training/no training. CFCCM used to rate content and quality. • Reliability of content rated excellent; mean kappa across all formulations, variables and coders = 0.86 (range 0.60-1.0). • Mean kappas for global CF quality, elaboration of four generic components, complexity, degree of inference and precision of language slightly lower; means were 0.75 between coders 1 & 2, 0.72 between coders 1 & 3, 0.81 between coders 2 & 3. Quality ratings higher for training group.	No	Kappa coefficient for inter-rater reliability of content categories and quality ratings of CF	Yes
Koerner (1997)	• Inter-rater reliability of ratings of relevance of individual CF items using ICC was poor for single judges (problem list: 0.32, goals: 0.31, hypothesized mechanism: 0.37) but fair when averaged across five judges (problem list: 0.66, goals: 0.66, hypothesized mechanism: 0.61). • Inter-rater reliability of ratings of relevance between expert (Persons) and other judges using ICC was poor to fair.	No	ICC for inter-rater reliability of ratings of relevance of CF items across raters and with expert judge	Yes

Authors (date)	Key findings	Use of systematized approach to formulation	Method of reliability assessment	Use of validated measure of reliability
Kuyken, Fothergill, Musa & Chadwick (2005)	• Percentage agreement with benchmark formulation was good to high for most categories (range 7-73%). • Good levels of agreement for relevant childhood experiences (critical father: 72%), core beliefs (I'm incompetent: 73%) and compensatory strategies (avoidance: 71%). • Low agreement for dysfunctional assumptions (e.g. I lack support and am isolated: 18%). • Good agreement on content categorisation for information contained in CF (range κ = 0.63-0.91).	Yes – J. Beck Case Conceptualisation Diagram	Percentage agreement with benchmark formulation and other raters for specific content of CF. Kappa coefficient for inter-rater reliability of content categorisation.	Yes
Muran, Samstag, Segal & Winston (1998)	• Ratings of relevance of self-scenarios and their components to the patient were adequate across three raters; mean overall rating of ICC (2,k) was 0.94 across scenarios (range 0.88-0.97). • Test-retest reliability of ratings for all self-scenarios (ICC:3,k) based on all patients over the 29 sessions ranged from 0.79 - 0.97.	Yes – self-scenario structure based on views of self/others/ world in context	ICC for inter-rater reliability of ratings of relevance of scenarios and their components, ICC for test-retest reliability of ratings	Yes

Authors (date)	Key findings	Use of systematized approach to formulation	Method of reliability assessment	Use of validated measure of reliability
Muran, Samstag, Ventur, Segal & Winston (2001)	• Interrater reliability for ratings of relevance of interpersonal scenarios and their components to patient was high across two raters (patient, therapist): mean ICC = 0.96. • Across three raters (interviewer, observer, patient), mean ICC = 0.87, across four raters (as before plus constructor), mean ICC = 0.92.	Yes – interpersonal scenario structure based on views of self/others/ world in context	ICC for inter-rater reliability of ratings of relevance of interpersonal scenarios and components	Yes
Muran & Segal (1992)	• Ratings of clinical relevance of self-scenarios high across both two (patient, therapist) and three (patient, therapist, observer) raters: mean ICC = 0.92 (two raters) and 0.93 (three raters). • ICCs for self-scenario components remained high across two raters: situation (0.92), affective (0.93), motoric (0.95), cognitive (0.90) and three raters: situation (0.92), affective (0.97), motoric (0.95), cognitive (0.88).	Yes – self-scenario structure based on views of self/others/ world in context	ICC for inter-rater reliability of ratings of relevance of self-scenarios	Yes
Muran, Segal & Samstag (1994)	• ICC (3,k) of ratings of relevance of self-scenarios to patients, across three raters (patient, therapist, third-party observer) showed adequate reliability across all cases for situation (0.92) cognitive (0.93) affective (0.90), motoric (0.91). • Mean ICC across all raters, parameters and cases = 0.92, range 0.79-0.99.	Yes – self-scenario structure based on views of self/others/ world in context	ICC for inter-rater reliability of ratings of relevance of self-scenarios	Yes
Nattrass, Kellett, Hardy & Ricketts (2015)	• Good inter-rater reliability for content and quality across every CF (multi-rater $\kappa = 0.64$. • Excellent overall inter-rater reliability for quality (ICC = 0.75). ICCs for individual quality ratings were 0.73 (coherence), 0.76 (precision of language), 0.78 (elaboration), 0.88 (systematic process) and 0.92 (complexity).	No	ICC for inter-rater reliability of content categories and quality ratings of CF	Yes

Authors (date)	Key findings	Use of systematized approach to formulation	Method of reliability assessment	Use of validated measure of reliability
Persons & Bertagnolli (1999)	• Mean percentage of problems correctly identified by individual therapists when compared to expert problem list was 67.46% (SD = 13%). • Schema ratings (patient's belief in this item) showed good inter-rater reliability averaged across five judges (ICC = mean 0.72, range 0.44-0.91). Poor inter-rater reliability of schema ratings for single judges (ICC = mean 0.37, range 0.13-0.66). • Analysis of variance demonstrated no increase in agreement on schema ratings when specific context rather than no context provided.	Yes – taught list of potential problem domains, multiple choice questionnaire listed adjectives describing possible self/other/world schemas with 0-10 scale and anchor points, context for schema ratings provided.	Percentage agreement with expert problem list. ICC for inter-rater reliability of ratings of schema relevance.	Yes
Persons, Mooney & Padesky (1995)	• Expert problem list: Case 1: Inter-rater reliability high for two problems (97.8%, 82.6%), poor for one (13.0%). Case 2: Inter-rater agreement was high for three problems (100%, 97.5%, 93.4%), moderate for two (67.4%, 71.7%). • Inter-rater reliability of ratings of centrality of proposed schemas to case and identification of dysfunctional attitudes found to be poor for single judges (ICC range = 0.07-0.55) but adequate across five judges for all ratings except one (dysfunctional attitudes - case 2, ICC = 0.27). ICC range excluding this rating mean = 0.83, range = 0.66-0.92.	Yes – multiple choice questionnaire listed adjectives describing possible self/other/world schemas with 0-10 scale.	Percentage agreement with expert problem list. ICC for inter-rater reliability of ratings of schema relevance	Yes

Note. *UCM* Underlying cognitive mechanism; *CF* Case formulation; *CFCCM* Case Formulation Content Coding Manual (Eells et al., 1998); *ICC* Intraclass correlation coefficient (Shrout & Fleiss, 1979).

Table 3

Study Evaluation Targets

Authors (date)	Inter-clinician agreement	Use of expert / benchmark	Evaluation of CF effect on outcome	Clinical utility (therapist)	Clinical utility (client)	Model integrity checked	Prescribed training	Formulation model	Therapist experience
Barris (1996)	√	√	√	X	X	X	X	Underlying cognitive mechanisms	
Chadwick, Williams & Mackenzie (2003)	X	X	√	√	√	X	X	CBT for psychosis	CBT Therapists: 6/15 yrs Trainee CPs: <1 yr
Dudley, Park, James & Dodgson (2010)	√	√	X	X	X	X	X	CBT Case Formulation	Yrs of clinical exp *m* = 8.5
Eells, Kendjelic & Lucas (1998)	√	X	X	X	X	X	X		
Eells, Lombart, Kendjelic, Turner & Lucas (2005)	√	X	X	X	X	X	X		Novices: <1500 hrs psychotherapy exp Experienced: 10 yrs+ therapy exp Expert: developed CF method / led 1+ workshops on CF / published on CF

Authors (date)	Inter-clinician agreement	Use of expert / benchmark	Evaluation of CF effect on outcome	Clinical utility (therapist)	Clinical utility (client)	Model integrity checked	Prescribed training	Formulation model	Therapist experience
Eells, Lombart, Salsman, Kendjelic, Schneiderman & Lucas (2011)	√	X	X	X	X	X	X		Novices: <1500 hrs psychotherapy exp Experienced: 10 yrs+ therapy exp Expert: developed CF method / led 1+ workshops on CF / published on CF
Flitcroft, James, Freeston & Wood-Mitchell (2007)	X	X	X	√	√	X	X	CBT	
Ghaderi (2006)	√	X	√	X	X	X	X	CBT individualised with logical functional analysis	
Haarhoff, Flett & Gibson (2011)	√	√	X	X	X	X	√	CBT	Novices: Recently completed post-grad dip in CBT
Hess (2000)	X	X	√	X	√	√	√	Cognitive Behavioural Case Formulation	8/15 yrs clinical exp

Authors (date)	Inter-clinician agreement	Use of expert / benchmark	Evaluation of CF effect on outcome	Clinical utility (therapist)	Clinical utility (client)	Model integrity checked	Prescribed training	Formulation model	Therapist experience
Kahlon, Neal & Patterson (2014)	X	X	X	X	√	X	X	CBT	
Kendjelic & Eells (2007)	√	X	X	X	X	X	√		Yrs exp with patients in mental health setting *m* = 2.9 (range 1 mth – 20 yrs)
Koerner (1997)	√	√	X	√	X	√	√	Cognitive Behavioural Case Formulation & Plan Formulation	9.5 yrs cognitive therapy exp
Kuyken, Fothergill, Musa & Chadwick (2005)	√	√	X	X	X	√	√	J.Beck Case Concept-ualisation Diagram	Wide range: min = basic knowledge & exp of CBT. Qualified therapists: Mean 7 years post-qual exp in clinical settings.
Morberg Pain, Chadwick & Musa (2008)	X	X	X	√	√	√	X	CBT Case Formulation	Training in CBT; 2 completed doctoral training, 2 in doctoral training, 1 in nurse training

Authors (date)	Inter-clinician agreement	Use of expert / benchmark	Evaluation of CF effect on outcome	Clinical utility (therapist)	Clinical utility (client)	Model integrity checked	Prescribed training	Formulation model	Therapist experience
Mumma (2004)	X	X	√	X	X	X	X	Idiographic Cognitive Schemas - CBT Case Formulation	
Mumma & Fluck (2016)	X	X	√	X	√	X	X	Functional analytical clinical case model	
Mumma & Mooney (2007)	X	X	X	X	X	X	X	Idiographic Cognitive Schemas - CBT Case Formulation	Clinician 1: expert Clinician 2: novice (clinical doctoral student)
Mumma & Smith (2001)	√	X	X	X	X	X	X	Cognitive-Behavioural-Interpersonal Scenarios	
Mumma, Marshall & Mauer (2018)	X	X	X	X	X	X	X	Functional analytical clinical case model	

Authors (date)	Inter-clinician agreement	Use of expert / benchmark	Evaluation of CF effect on outcome	Clinical utility (therapist)	Clinical utility (client)	Model integrity checked	Prescribed training	Formulation model	Therapist experience
Muran, Samstag, Segal & Winston (1998)	√	X	√	X	X	X	√	Interpersonal scenarios	Mean 9.5 yrs of clinical exp
Muran, Samstag, Ventur, Segal & Winston (2001)	√	X	√	X	X	X	X	Interpersonal scenarios	
Muran & Segal (1992)	√	X	√	X	X	X	X	Self-scenarios	6 yrs clinical exp (range 3-9 yrs)
Muran, Segal & Samstag (1994)	√	X	√	X	X	X	X	Self-scenarios	Mean years of exp: 3.1 (SD = 1.3)
Nattrass, Kellett, Hardy & Ricketts (2015)	√	X	√	X	X	√	√	CBT	2-10 yrs post-qualification CBT exp
Persons & Bertagnolli (1999)	√	√	X	X	X	X	√	Cognitive Behavioural Case Formulation	Mean = 10 years
Persons, Bostrom & Bertagnolli (1999)	X	X	√	X	X	X	√	Cognitive Behavioural Case Formulation	10 years of clinical exp

Authors (date)	Inter-clinician agreement	Use of expert / benchmark	Evaluation of CF effect on outcome	Clinical utility (therapist)	Clinical utility (client)	Model integrity checked	Prescribed training	Formulation model	Therapist experience
Persons, Mooney & Padesky (1995)	√	√	X	X	X	X	X	Cognitive Behavioural Case Formulation	
Persons, Roberts, Zalecki & Brechwald (2006)	X	X	√	X	X	X	X	Cognitive Behavioural Case Formulation	Lead therapist (author) 20 years, 3 additional therapists: 2-6 years
Redhead, Johnstone & Nightingale (2015)	X	X	X	X	√	X	X	CBT	

Note. Blank entry in table indicates relevant information was not provided by study; *CF* Case Formulation; References for specific formulation models given when this was noted in the study: *Underlying cognitive mechanism* (Persons, 1989*); CBT case formulation* (Beck, 1976); *Cognitive Behavioural Case Formulation* (Persons, 1989); *Plan Formulation* (Rosenberg, Silberschatz, Curtis, Sampson & Weiss, 1986; *J. Beck Case Conceptualization Diagram* (Beck, 1995); *Idiographic Cognitive Schemas – CBT Case Formulation* (Beck, 1995); *Functional analytical clinical case model* (Haynes et al., 2011); *Cognitive Behavioural Interpersonal Scenarios* (Mumma & Smith, 2001); *Interpersonal scenarios* (Muran, 1993); *Self-scenarios* (Muran & Segal, 1991).

Evaluation of Validity

Details of validity assessment within included studies are listed in Tables 4 and 5, with additional information in Table 3. Twenty-eight studies included evaluation of the validity of cognitive case formulation (93.33%); of these, sixteen considered both reliability and validity (53.33% of all included studies) and twelve validity alone (40.00%).

Twenty-one studies obtained material for analysis by asking clinicians to generate formulations in response to stimulus material, which included live clinical interviews (*n* = 10), videos of clinical interviews (*n* = 4), audiotapes of clinical interviews (*n* = 2), written vignettes (*n* = 3), and session transcripts (*n* = 1). Four studies included interviews reflecting on the experience of receiving a formulation, three studies examined existing clinical material (intake reports and formulations), one considered statements about what might be important in cognitive therapy and one study did not state the source material for formulation.

The definition of validity in relation to cognitive formulation is not straightforward. In psychometric terms, validity encompasses a range of concepts, including content, construct, criterion, discriminant and predictive validity. Mumma (2011) has described content validity as the extent to which a formulation includes and represents target and causal variables and the relationships between them. Construct validity (including convergent and discriminant validity) is the extent to which a formulation of someone's problems captures theoretically important aspects such as symptom severity and provides a coherent account of problems, criterion validity describes the extent to which formulation matches expert conceptualisation and predictive validity is a measure of the extent to which formulations predict outcomes and changes over the course of therapy (Bucci et al., 2016).

Operationalising these concepts in terms of cognitive case formulation is hard to achieve directly and included studies assessed formulation validity in a range of ways, with some studies using more than one method. The most common approach was to examine outcome scores for formulation-driven treatment, under the assumption that a valid formulation was likely to

contribute to improved outcomes (*n* = 9; Barris, 1996; Chadwick et al., 2003; Ghaderi, 2006; Hess, 2000; Muran & Segal, 1992; Muran et al., 1994; Nattrass et al., 2015; Persons et al., 1999; Persons et al., 2006). Six studies, encompassing the three qualitative studies in the review, considered the clinical utility of formulation as a measure of validity, from either or both clinicians' and clients' perspectives (Chadwick et al., 2003; Flitcroft et al., 2007; Kahlon et al., 2014; Koerner, 1997; Morberg Pain et al., 2008; Redhead et al., 2015). Seven studies used often novel combinations of empirical data collection and statistical analysis to evaluate the validity of formulation in a more psychometric way, often with a high degree of collaborative input from clients (Mumma, 2004; Mumma & Fluck, 2016; Mumma & Mooney, 2007; Mumma & Smith, 2001; Mumma et al., 2018; Muran et al., 1998; Muran et al., 2001). Finally, eight studies included an assessment of formulation quality (Dudley et al., 2010; Eells et al., 1998, Eells et al., 2005; Eells et al, 2011; Kendjelic & Eells, 2007; Nattrass et al., 2015). The relationships between formulation quality and validity are of interest to clinicians, given evidence that higher quality formulations have been found to contribute to better outcomes (Crits-Christoph, Cooper & Luborsky, 1988). These studies examined factors which may influence quality, as well as the relationship between quality and outcome, aiming to increase understanding of factors likely to relate to formulation validity.

Following data extraction, the key findings of these studies were, first, that cognitive formulation-driven treatment can be associated with statistically and clinically significant changes in outcomes; three studies (Nattrass et al., 2015; Persons et al., 1999; Persons et al., 2006) found large pre-post therapy reductions in symptoms, with Persons et al. (2006) also finding that half or more of patients in the study met the definition for clinical recovery from anxiety and/or depression. Furthermore, when compared with manualised treatment in a randomised controlled trial, results from Ghaderi (2006) suggested that individualised treatment produced better outcomes. However, this was not supported by Barris (1996), which found no difference in outcome between manualised and individualised treatment. Two studies (Muran et al, 1998; Muran et al., 2001) found correspondences between trends in clients' ratings of key dimensions of self-scenario formulations

over time and good/bad outcomes to therapy, suggesting that these formulations were capturing factors related to outcome. However, Hess (2000) found no connection between higher ratings for the accurate identification of a core issue and outcome measures, including working alliance, perceived empathy and symptom distress. Overall, explorations of the relationship between formulation and outcome which attempted to develop understanding of validity produced mixed and inconclusive results.

Assessment of clinical utility as a proxy for validity produced similarly mixed results. Four studies used qualitative approaches to explore experiences of formulation, with three using thematic analysis to evaluate findings and one (Morberg Pain et al., 2008) using content analysis. Chadwick et al. (2003), which explored both therapists' and clients' experiences, found clinicians' responses very positive, including feeling validated, hopeful and having a sense that sharing formulation had enhanced understanding. Clients' responses within the same study, as well as in Kahlon et al. (2014), Morberg Pain et al. (2008) and Redhead et al. (2015) expressed much greater ambivalence about the utility of formulation, including negative emotional reactions to receiving them, although when formulation was perceived to be accurate, this appeared to enhance a sense of being understood. Collaboration in producing a formulation appeared important in enhancing the clients' sense of a formulation being accurate (Kahlon et al., 2014). These differing views about the clinical utility of formulation were supported by Flitcroft et al. (2007) which used a Q-sort methodology to understand clinicians' perspectives on the key function of formulation; little consensus was found, with three very different factors ('state', 'function and process', and 'trait/historical aspects') all perceived to be important.

Seven studies employed statistical analysis of data related to a client's symptoms and distress, often collected longitudinally as well as during an initial formulation process, in order to evaluate the content, construct and/or predictive validity of formulations. These studies used very small samples, often just a single client as a test of process (*n* = 4; Mumma, 2004; Mumma & Fluck, 2016; Mumma et al., 2018; Muran et al; 2001) but proved the utility of a statistically-driven

approach to test a hypothesis about a client's difficulties and produce evidence for its validity in these terms. In Mumma and Mooney (2007), two clinicians independently produced proposed idiographic cognitive schemas for the same client, based on viewing a clinical interview; the construct and predictive validity of each model was then tested using data from daily questionnaires and the most valid model identified. While their process differed in focusing more on initial formulation, both Muran et al. (1998) and Muran et al. (2001) analysed multiple ratings of the same self-scenario by the client, therapist and observer to establish the criterion validity of those formulations, as well as using clients' repeated ratings of key dimensions of those self-scenarios over time to establish their predictive validity. Although these processes only allow validity to be established retrospectively, following complex statistical analysis, these studies suggest potential directions for exploring the clinical use of simpler versions of these methods.

Finally, included studies' evaluations of formulation quality (*n* = 8) failed to clarify its potential as evidence for validity and did not clearly support formulation quality as a factor in therapy outcome. Eells et al. (1998) found the quality of formulations included in their study to be low; formulations were generally descriptive with little inferential information (mean score for inferential content on 1-5 scale was 1.80, SD 0.77). Haarhoff et al. (2011), in a study using four separate measures of content and quality with novice clinicians, found formulation quality to be moderate overall. While some aspects were consistently well explained (developmental factors, psychological mechanisms), others were often omitted (socio-cultural factors, therapeutic relationship), with the poorest quality formulations imprecise and disorganised. Quality assessment of formulation in Kuyken et al. (2005) found only 44.2% of formulations 'good enough', although noted that experience, professional qualification and accreditation were associated with higher quality formulations. Other studies also evaluated factors with the potential to improve formulation quality including experience (Dudley et al., 2010; Eells et al., 2005, Eells et al., 2011) and training (Kendjelic & Eells, 2007), with Dudley producing evidence for the beneficial effect of experience with CBT cases on inclusion of inferential material, both studies by Eells and colleagues suggesting that

expertise, rather than experience, improved formulation quality and Kendjelic and Eells (2007) obtaining a large effect size for the impact of training on formulation quality. While these studies indicated that formulation quality could be improved, the purpose of this improvement remained unclear from Nattrass et al. (2015), which concluded that, although formulation-based treatment could produce good outcomes, those outcomes appeared independent of formulation quality, with no significant correlation found between formulation quality and outcome.

Table 4

Evaluation of Validity - Quantitative Studies

Authors (date)	Key findings	Format, intensity & duration of intervention	Method of validity assessment	Validated measures used?
Barris (1996)	• Randomised controlled trial; no difference in efficacy found between individualised and standardised treatment conditions overall. • Within individualised group, patients with sociotropic underlying cognitive mechanism did better than autonomous patients on symptomatic and cognitive measures (post-treatment scores for autonomous patients higher on BDI ($F(2.25) = 4.97$, $p = 0.03$) and lower on SPB total score ($F(2.25) = 3.87$, p+ 0.06). • No clear relation between CF and better outcome.	Both conditions' weekly program: groups, individual work, homework. Average duration of inpatient stay: Individualised group = 14.7 days; standard = 13.2 days.	Changes in outcome measures pre-and post-treatment (BDI, ABS-2, SPB)	Yes
Chadwick, Williams & Mackenzie (2003)	• Experiment 1: Impact of CF on anxiety or depression was non-significant using Friedman's ANOVA (Dep: Chi Squared = 6.78, df = 3, non-significant: Anxiety: Chi Squared = 3.32, df = 3, non-significant). Client scores on HAq-P showed significant increase between time points T1 and T3 ($T = -2.12$, $p < 0.05$) and T1 and T4 ($T = -2.25$, $p < 0.05$); insufficient to support CF having significant impact for clients. • Experiment 2: No significant impact of CF on distress, secondary delusions & negative beliefs about self. Beliefs did weaken during subsequent direct challenge, suggesting CF alone ineffective. • E1 & E2: No impact of CF on distress from combined results (Depression. Chi squared=1.79, df=2, non-significant: Anxiety. Chi squared=0.094, df=2, non-significant).	None	Change in outcome measures during sharing of CF (HADS, HAq, BAVQ-R, PSYRATS)	Yes

Authors (date)	Key findings	Format, intensity & duration of intervention	Method of validity assessment	Validated measures used?
Dudley, Park, James & Dodgson (2010)	• Positive association between mean score for 'number of CBT cases seen' and number of inferential elements included in CF and (β = 0.28, *t* = 2.17, p < 0.05). • Therapist experience positively associated with higher mean score (β = 0.42, t = 2.85, p < 0.01). • 'Number of cases of psychosis seen' inversely related to mean score for CF (β = -0.37, *t* = -2.38, p < 0.05).	None	Associations between therapist experience and agreement about content of CF	Yes
Eells, Kendjelic & Lucas (1998)	• Quality ratings of CFs: 55.4% did not include presentation of a mechanism. • Rating of complexity showed formulations were relatively simple, with little evidence of integration of information (*m* = 2.05, SD = 0.94). • Ratings of inference indicated primarily descriptive formulations with little inferential information (*m* = 1.80, SD = 0.77; formulations were moderately precise in their language (*m* = 2.57, SD = 0.93). All scales 1-5 (5 best).	None	Ratings of CF quality using CFCCM	Yes
Eells, Lombart, Kendjelic, Turner & Lucas (2005)	• Experts' CFs rated as more comprehensive, complex, elaborated and fitting in terms of treatment plan, precise and systematic than those of experienced or novice clinicians. • Large effect size on measure of overall quality when expert vs. experienced clinicians compared (0.49). • Novice clinicians CFs rated higher than experienced clinicians' CFs overall.	None	Ratings of CF content and quality using CFCCM comparing novice, experienced, expert clinicians	Yes

Authors (date)	Key findings	Format, intensity & duration of intervention	Method of validity assessment	Validated measures used?
Eells, Lombart, Salsman, Kendjelic, Schneiderman & Lucas (2011)	• Experts: more descriptive information (*r* = 0.33), more focus on diagnosis (*r* = 0.24), more inferential information (*r* = 0.35) more extensive treatment planning (*r* = .32) than non-experts; effect sizes small to moderate. • Experts' treatment plans: more detailed (*r* = 0.34) more symptoms and psychological mechanisms inferred (both *r* = 0.25).	None	Ratings of CF content and quality using CFCCM comparing novice, experienced, expert clinicians	Yes

Flitcroft, James, Freeston & Wood-Mitchell (2007)	• Q-sort methodology applied to 86 statements. Three factors identified with eigenvalues above 1, and accounting for 49% of variance:
	  ○ Factor A: a 'state' factor, focusing on 'here and now' 24%
	  ○ Factor B: 'function and process', 18%
	  ○ Factor C, 'trait/historical aspects', 7%
	• Three factors represented very different perspectives; these differences may account for low inter-rater reliability of CFs.
		None	Use of Q-sort to capture clinicians' views of most important aspects of case formulation	Yes

Ghaderi (2006)	• Randomised controlled trial comparing outcomes for individualised vs. standard treatment for bulimia nervosa. • Outcomes on some EDE subscales better for individualised over manualised treatment, moderate to large effect sizes (partial Eta squared): number of weeks of abstinence from objective bulimic episodes (0.15), eating concerns (0.11) and body shape dissatisfaction (on BSQ) (0.08). • Significant group interaction was observed for excessive exercising (0.15) and compensatory behaviour (0.10). • 92% in individualised group vs. 69% in standardised group responded to treatment.	19 x weekly CBT	Comparison of outcomes for individualised vs. standard treatment groups (EDE interview & questionnaire, BSQ, RSE, PSS, HSCL-90, BDI, WCQ)	Yes

Authors (date)	Key findings	Format, intensity & duration of intervention	Method of validity assessment	Validated measures used?
Haarhoff, Flett & Gibson (2011)	• Quality assessment of formulation using four measures found that quality was moderate (mean rating 3 (0-4 scale) on QCTCFRS, rating on CBT CC between 17.2-18.8/40 for the four vignettes). • Consistent strengths: inclusion of development factors, psychological mechanisms. • Consistent weaknesses: lacked socio-cultural or biological factors, therapeutic relationship, lack of specific information, purely descriptive.	None	Ratings of CF content and quality using comprehensiveness scale from CFCCM, QCTCFRS, CBT CC and benchmark formulations	Yes
Hess (2000)	• Accuracy of identification of 'core underlying mechanism' within client's CF as driver of symptom change assessed. • Validity/accuracy of core issue rated but interrater reliabilities of ratings poor (0.11 for one pair and 0.05 for the other pair). • No link between higher ratings of core issue accuracy and measures (e.g. working alliance, perceived empathy, symptom distress). • Validity of core issue in relation to client responses showed mixed; range from uncertain about accuracy of CI to agreeing but later showing ambivalence.	Weekly CBT (mean *n* = 12.8)	Change in outcome scores correlated with sharing of 'core underlying cognitive mechanism' evaluated using OQ-45, SCS, WAI, ES, SEQ	Yes
Kendjelic & Eells (2007)	• Randomised controlled trial comparing clinician formulations following CF training/no training. • Quality ratings of CFs showed large mean effect size (1.12), indicating average clinician in training group produced better quality formulations than 86% of those in control (no training) group.	None	Ratings of CF content and quality using CFCCM	Yes

Authors (date)	Key findings	Format, intensity & duration of intervention	Method of validity assessment	Validated measures used?
Koerner (1997)	• Expert rated individual CF items' relevance across cases on Likert scale 0-4 (mean: problem list: 2.9; goals: 2.8; hypothesized mechanism: 2.7). No significant differences from formulation or reliability judges' ratings.	None	Ratings of relevance of CF items by expert / formulation / reliability judges.	No
Kuyken, Fothergill, Musa & Chadwick (2005)	• Quality assessment found good to high agreement with benchmark formulation. • QCCFRS: Formulations rated 'very poor' (22.1%), poor (33.6%), 'good enough' (34.5%), good (9.7%). Only 44.2% at least 'good enough'. • Considerable variation in how parsimoniously, coherently and meaningfully formulation fitted together. • Number of years of post-qualified experience associated with high quality formulation. • Professional qualification and BABCP accreditation appeared to lead to incremental improvement in formulation quality.	None	Ratings of quality using QCCFRS and comparison with benchmark formulation.	No – developed for this study
Mumma (2004)	• Idiographic Cognitive Schemas found to explain an average of 29% of the daily variability in symptom/distress ratings (R^2 = 0.20 to 0.44). • This was measured on an Individualised Daily Questionnaire, collected once daily for 90 days, and was comparable between average shrunken R^2 = 0.24 for ICS predictors and R^2 = 0.25 for nomothetic cognition predictors. • Incremental validity: Idiographic Cognitive Schema factor scores predicted mean additional 6% of variability above that predicted by nomothetic cognitions (R^2 = 0.0 - 0.17).	Weekly CBT	Analysis of convergent & discriminant validity using confirmatory dynamic factor analysis of multivariate time-series data collected daily; analysis of dynamic structure of Idiographic Cognitive Schemas using dynamic time series regression.	Yes

Authors (date)	Key findings	Format, intensity & duration of intervention	Method of validity assessment	Validated measures used?
Mumma & Fluck (2016)	• Empirically testing individual cognition-distress hypotheses within the CF developed by clinician & client was important and allowed identification of valid/invalid hypotheses.	CBT	Correlational analysis of data collected through daily surveys of symptoms identified as key by client	Yes
Mumma & Smith (2001)	• Good convergent validity for the CBIS sets for the same patient generated by two different clinicians: convergent validity coefficients for eight patient-by-formulator centroids relatively small (*m* = 0.469). • Good discriminant validity on one discriminant validity coefficient, with distance between CBIS sets for different patients for CBIS sets generated by same clinician producing large scores (*m* = 1.000, SD - 0.375). • Good discriminant validity found for CBIS sets for different patients developed by different clinicians; large scores (*m* = 1.043, SD = .370). • CBIS sets demonstrated satisfactory convergence across formulators for same patient and divergence between patients for same or different formulator using 3-factor solution, with the dimensions anxiety, depression and interpersonal functioning. Good within-case convergent and discriminant validity also demonstrated.	None	Analysis of data collected on CBISs for two clients using individual daily questionnaires; convergent & discriminant validity assessed using multi-trait multi-method matrix	No

Authors (date)	Key findings	Format, intensity & duration of intervention	Method of validity assessment	Validated measures used?
Mumma, Marshall & Mauer (2018)	• Good item level construct validity for distress and cognition scales (a = 0.88 - 0.97). • Acceptable item level discriminant validity (mean adjusted-item total correlation of items: Depression – 0.85, Worry & Avoidance – 0.77. • Scale-level discriminant validity analysis showed good validity following detrending (r = 0.21). • Functional relation testing using bivariate regression analysis showed that 55% of hypothesized concurrent relations between triggers and distress were supported (6/11).	Not stated	Analysis of item-level construct and discriminant validity, scale-level discriminant validity and testing of functional relations using correlational analysis, regression and structural equation modelling of causal and distress scales	Yes
Muran, Samstag, Segal & Winston (1998)	• Support for validity of content of self-scenarios found: corrected item-total correlations of ratings of relevance of self-scenarios with patient's ratings as predictor (total scenarios *m* = 0.8). • Three out of six cases treated using this method showed 'good outcome' based on scores for reliable change (good outcome = RC > 1.96) on the global severity index of the SCL-90R and mean index of IIP. • Time-series analysis of self-scenario parameters 'Frequency' and 'Self-Efficacy' showed improvements in line with nomothetic measures.	30 x individual psychotherapy	Criterion validity: calculated corrected item-total correlations using patient ratings of the relevance of self-scenarios as predictor. Predictive validity: time-series analysis of 2 x self-scenario parameters	Yes

Authors (date)	Key findings	Format, intensity & duration of intervention	Method of validity assessment	Validated measures used?
Muran, Samstag, Ventur, Segal & Winston (2001)	• Criterion validity of content of self-scenarios: high ICC (2,k) for ratings of relevance of self-scenarios with patient's ratings as criterion (total scenarios = 0.87). • Evidence for efficacy of formulation-drive treatment: ratings for self-scenario parameters 'Frequency' (of this scenario occurring) and 'Self-Sufficiency' showed increased occurrence of best-case scenario and converse downward trend for worst-case scenarios plus increase in sense of self-sufficiency during formulation-driven therapy.	40 x cognitive / interpersonal therapy	Criterion validity: calculated corrected item-total correlations using patient ratings of the relevance of self-scenarios as criterion. Predictive validity: time-series analysis of 2 x self-scenario parameters	Yes
Muran & Segal (1992)	• Visual analysis of data showed complementary modest to substantial decreases in ratings for negative aspects of self-scenarios and substantial increases in positive aspects with improved scores on the SCL-90 over the course of therapy (T-score on Global Severity Index improved five points after 10 sessions, nine points after 20 sessions, score on depression subscale improved 2 points after 10 sessions, 16 points after 20 sessions). • GAS rating improved from moderate to mild symptomology by treatment end.	20 x weekly cognitive therapy	Changes in clinical outcome scores corresponding to changes in ratings of aspects of self-scenarios.	Yes
Muran, Segal & Samstag (1994)	• Composite score based on Global Severity Index of SCL-90R and sum total score on DAS used to assess reliable change for clients over course of therapy. Only two out of eight cases showed significant change by end of formulation-driven therapy, based on this index.	20 x weekly cognitive therapy	Correspondence between trend in scores for aspects of self-scenario with changes in clinical outcome scores over course of therapy.	Yes

Authors (date)	Key findings	Format, intensity & duration of intervention	Method of validity assessment	Validated measures used?
Nattrass, Kellett, Hardy & Ricketts (2015)	• Pre-post outcomes showed large effect size (Cohen's *d* = .89) on the Y-BOCS, with a reliable and clinically significant change in OCD symptoms between the assessment and end of formulation-driven treatment. • No significant correlation between formulation quality and outcome during formulation or post-formulation phases on Y-BOCS (mean r = 0.36 formulation phase, 0.37 post formulation), CORE-SF (mean r = -0.24 formulation, 0.05 post formulation), or ARM-12 (mean r = 0.08 formulation phase, 0.05 post formulation). • Significant reduction in psychological distress on CORE-SR between assessment (*m* = 18.23, 95% CI [15.48, 20.99] and formulation phases (*m* = 17.20, 95% CI [13.61, 20.80] and post-formulation (*m* = 14.81, 95% CI [11.55, 18.08]. • Therapeutic alliance at end post-formulation phase (*m* = 75.41, 95% CI [72.34, 78.49] significantly higher than during assessment (*m* = 70.28, 95% CI [67.07, 73.48].	CBT	Analysis of correlations between formulation quality and symptomatic change, including psychological distress and therapeutic alliance.	Yes
Persons, Bostrom & Bertagnolli (1999)	• Randomised controlled trial: Paired *t* tests showed patients in private practice (both treated with formulation-driven cognitive therapy and formulation-driven cognitive therapy and pharmacology) showed a statistically significant decrease in BDI scores [*t*(44) = 9.37, p=.0001]. • Patients treated with formulation-driven cognitive therapy alone (*n* = 27, d = 1.44) and formulation-driven CBT with pharmacology (*n*=18, *d*=1.41) both showed large pre-post effect sizes for change in depressive symptoms on the BDI. These were comparable to RCT outcomes.	Weekly CBT (mean *n* = 34.8)	Changes in BDI scores pre-and post-treatment	Yes

Authors (date)	Key findings	Format, intensity & duration of intervention	Method of validity assessment	Validated measures used?
Persons, Roberts, Zalecki & Brechwald (2006)	• Patients who received case-formulation driven CBT showed statistically significant reductions with a large effect size in symptoms of depression and anxiety over the course of treatment (depression: BDI effect size = 1.33, anxiety: BURNS AI effect size = 0.98). • Based on improvement being a 50% reduction in symptoms 55.3% (BDI) and 46.4% (BURNS-AI) of patients showed improvement. 50% of patients recovered from depression (BDI score 9 or less); 57% recovered from anxiety (BURNS AI score lower than mean of control sample).	Formulation-driven CBT (mean *n* = 18) plus pharmacotherapy, 12-step groups, couple therapy	Changes in outcome measures pre-and post-treatment (BDI, BURNS AI)	Yes

Note. *CF* Case Formulation; *BDI* Beck Depression Inventory (Beck & Steer, 1987); *SPB* Survey of Personal Beliefs (Demaria et al., 1989); *ABS-2* Attitudes and Beliefs Scale-2 (DiGiuseppe et al*., 1988); HADS* Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983); *HAq* Helping Alliance Questionnaire (Alexander & Luborsky, 1986); *BAVQ-R* Beliefs About Voices Questionnaire (Chadwick et al., 2000); *PSYRATS* Psychotic Symptoms Rating Scales (Haddock et al., 1999); *CFCCM* Case Formulation Content Coding Manual (Eells et al., 1998); *CBT* Cognitive Behavioural Therapy; EDE Eating Disorders Examination (Cooper & Fairburn, 1987); *BSQ* Body Shape Questionnaire (Cooper et al., 1987); *RSE* Rosenberg Self Esteem Scale (Rosenberg, 1979); *PSS* Perceived Social Support (Procidano & Heller, 1983); *HSCL-90* Hopkins Symptom Checklist-90 (Lipman et al., 1979); *WCQ* Ways of Coping Questionnaire (Folkman & Lazarus, 1988); *QCTCFRS* Quality of Cognitive Therapy Case Formulation Rating Scale (Fothergill & Kuyken, 2002); *CBT CC* Cognitive Behavioural Therapy Case Formulation rating scale (Haarhoff, 2008); *OQ-45* Outcome Questionnaire (Lambert et al., 1996); *SCS* Stages of Change Scales (McConnaughy et al., 1989); *WAI* Working Alliance Inventory (Horvath & Greenberg, 1989); *ES* Empathy Scale (Persons & Burns, 1985); *SEQ* Session Evaluation Questionnaire (Stiles, 1980); *BABCP* British Association for Behavioural & Cognitive Psychotherapies*; CBIS* Cognitive Behavioural Interpersonal Scenarios (Mumma & Smith, 2001); *SCL-90R* Symptom Checklist 90-Revised (Derogatis, 1983); *IIP* Inventory of Interpersonal Problems (Horowitz et al., 1988); *ICC* Intraclass correlation coefficient (Shrout & Fleiss, 1979); *SCL-90* Symptom Checklist 90 (Derogatis, 1977), *GAS* Global Assessment Scale (Endicott et al., 1976); *Y-BOCS* Yale-Brown Obsessive Compulsive Scale (Goodman et al., 1989); *OCD* Obsessive-Compulsive Disorder; *CORE-SF* (Clinical Outcomes in Routine Evaluation-Short Form (Evans et al., 2002); *ARM-12* Agnew-Davies Relationship Measure (Cahill et al., 2012); *CI* Confidence Interval; BURNS *AI* Burns Anxiety Inventory (Burns, 1998).

Table 5

Evaluation of Validity – Qualitative Studies

Authors (date)	Key findings	Analytical approach	Themes relevant to perception/experience of formulation validity	Author's conclusions
Chadwick, Williams & Mackenzie (2003)	Responses to open-ended questions: • Clients reported CF helped understanding of own probs (*n* = 9/11) felt reassured, encouraged, more optimistic (6), felt understood (3), negative reaction including felt upset, sad, worried, overwhelmed by probs (6), no emotional impact (2), found CF complicated (2). • Therapists reported more positive responses than clients, including feeling validated if client endorsed CF, hopeful, sense sharing CF had enhanced alliance, increased understanding of problems and enhanced model adherence.	Semi-structured interviews with open questions about relevance & helpfulness of CF and reactions to it. Small *n* of responses analysed using thematic analysis.	Both therapists and clients felt CF helped understanding, but clients reported much more mixed emotional reactions.	Different perceptions of CF by therapists and clients; therapist more positive. Therapist perception of beneficial effect on alliance not supported by HAq data.

Authors (date)	Key findings	Analytical approach	Themes relevant to perception/experience of formulation validity	Author's conclusions
Kahlon, Neal & Patterson (2014)	• One client commented that formulation felt accurate, they 'already knew' elements. Another client: formulation was inaccurate, creating difficulties with connecting to it. • Collaborative development of CF appeared to enhance client's perception of being understood and CF being accurate.	Thematic analysis	Four superordinate themes, three relevant: • The development of the formulation - from coming to my own conclusions to something the therapist developed • From negative to mixed feelings: emotional reactions to the formulation during the therapeutic process • A new journey: towards making a new sense of oneself	Collaborative development of CF appeared to enhance client's perception of being understood and CF being accurate
Morberg Pain, Chadwick & Musa (2008)	• 12% of coding units in qualitative analysis on theme of CF being 'accurate' and reflecting experience.	Content analysis using inductive coding	Perceptions of CF accuracy: • Accurate: 12% of coding units, mentioned in 7/13 transcripts; • Inaccurate/incomplete :4/13 transcripts • Accuracy uncertain: 1/13 transcripts)	Varied perceptions of CF accuracy

Authors (date)	Key findings	Analytical approach	Themes relevant to perception/experience of formulation validity	Author's conclusions
Redhead, Johnstone & Nightingale (2015)	• Four themes and 10 subthemes identified from analysis. Clients reported the perceived accuracy of the formulation facilitated feeling understood.	Inductive thematic analysis	Main themes: • Formulation helps me understand my difficulties • Formulation leads to feeling understood and accepted • Formulation leads to an emotional shift • Formulation enables me to move forward	Formulation best shared collaboratively; distress may be part of therapeutic process but should be managed sensitively.

Note. *HAq* Helping Alliance Questionnaire (Alexander & Luborsky, 1986).

Evaluation of Formulation Quality

Clarifying the relationships between reliability, validity and quality of formulation in order to identify their contributions to formulation utility has become a focus of recent literature concerning cognitive case formulation. Eight studies included in this review assessed the quality of case formulation (Dudley et al., 2010; Eells et al., 1998; Eells et al., 2005; Eells et al., 2011; Haarhoff et al., 2011; Kendjelic & Eells, 2007; Kuyken et al., 2005; Nattrass et al., 2015), with their findings summarised earlier.

The intention of this review to elucidate how the quality of case formulation may influence reliability and validity required examining the quality of the case formulations within all included studies. As most studies had not attempted this, criteria for an independent quality assessment were needed. Criteria were agreed with two senior academic researchers prior to this review taking place and are listed in Table 6, which also includes the ratings for all studies according to those criteria. These criteria overlap with many of the criteria used in existing formulation quality assessment tools, such as the quality criteria for the CFCCM (Eells et al., 1998) and the Quality of Cognitive Case Formulation Rating Scale (QRS; Kuyken et al., 2005). Both these tools were used in included studies which carried out quality assessment, along with the Cognitive Behavioural Therapy Case Conceptualisation Rating Scale (CBT CC; Haarhoff et al., 2011) and expert formulations, used as a benchmark for quality (CFCCM *n* = 5, QRS *n* = 2, benchmark *n* = 3, CBT CC *n* = 1; studies used more than one measure). The key criteria for the CFCCM are quality rating of the formulation as a whole, for each major subcategory (symptoms, predisposing life events, precipitating factors and mechanisms), and for the comprehensiveness and complexity of the formulation, degree of elaboration, degree of inference used, coherence, precision of language, elaboration of treatment plan, goodness of fit between formulation and treatment plan and systematic process (Eells et al., 1998; Eells et al., 2005).

Quality assessment using our criteria was significantly limited by the absence of raw formulation data from all studies, with the exception of brief excerpts. This meant that ‘no

information' was the most frequent rating given. Where ratings were given, they were based on the best information from studies, although this necessitated some assumptions. Where studies stated that independent raters had cross-checked formulations, a positive rating was given for 'material correctly interpreted' unless disagreement on interpretation was indicated. Where client collaboration was indicated, a positive rating was given for 'relevant'. Given the focus of some studies on specific aspects of formulation (e.g. self-scenarios in Muran et al., 1992) these formulations did not aim to be comprehensive and so this criterion was marked not applicable. Some aspects, such as 'comprehensive' and 'objective,' were particularly difficult to rate without raw formulation data, which meant that positive ratings were low for these criteria. Where a criterion could be reported, it was noted as present (√) or absent (X).

Given the arbitrary impact of the level of formulation detail reported on ratings of quality for a particular study, total ratings for formulation quality per study were not calculated as comparisons would be invalid. Within individual categories, with the caveat that there may also be some impact of what was reported, the criterion with the most positive ratings was 'provides basis for intervention' (26), followed by 'clear relationships' (16) and 'relevant' (15). Table 6 summarises the findings of this evaluation.

Table 6

Evaluation of Formulation Quality

Authors (date)	Comprehensive	Relevant	Objective	Evidence-based	Material correctly interpreted	Clear relationships	Clear causal relationships	Explains provenance where necessary	Moderators clearly described	Coherent	Mechs of change clearly described	Provides basis for intervention	Provides rationale for prioritisation
Barris (1996)	X	√	X	X	√	√	√	X	X	NI	X	√	X
Chadwick, Williams & Mackenzie (2003)	NI	√	NI	NI	√	NI	NI	Ni	Ni	NI	NI	√	NI
Dudley, Park, James & Dodgson (2010)	X	NI	NI	√	NI	√	X	X	X	NI	X	√	X
Eells, Kendjelic & Lucas (1998)	X	NI	NI	NI	NI	NI	X	X	X	X	X	√	X
Eells, Lombart, Kendjelic, Turner & Lucas (2005)	NI	NI	NI	√	NI	√	√	√	NI	√	√	√	NI

Authors (date)	Comprehensive	Relevant	Objective	Evidence-based	Material correctly interpreted	Clear relationships	Clear causal relationships	Explains provenance where necessary	Moderators clearly described	Coherent	Mechs of change clearly described	Provides basis for intervention	Provides rationale for prioritisation
Eells, Lombart, Salsman, Kendjelic, Schneiderman & Lucas (2011)	X	NI	NI	√	NI	NI	NI	√	NI	NI	NI	NI	NI
Flitcroft, James, Freeston & Wood-Mitchell (2007)	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
Ghaderi (2006)	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
Haarhoff, Flett & Gibson (2011)	X	NI	NI	X	NI	√	√	√	X	NI	NI	√	√
Hess (2000)	NI	√	NI	NI	√	NI	NI	√	NI	NI	NI	√	NI
Kahlon, Neal & Patterson (2014)	NI	√	NI	NI	NI	NI	NI	NI	NI	NI	NI	√	X
Kendjelic & Eells (2007)	√	NI	NI	NI	NI	√	√	√	X	NI	√	√	X
Koerner (1997)	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	√	NI

Authors (date)	Comprehensive	Relevant	Objective	Evidence-based	Material correctly interpreted	Clear relationships	Clear causal relationships	Explains provenance where necessary	Moderators clearly described	Coherent	Mechs of change clearly described	Provides basis for intervention	Provides rationale for prioritisation
Kuyken, Fothergill, Musa & Chadwick (2005)	NI	NI	NI	√	NI	√	√	√	NI	√/X	√	√	NI
Morberg Pain, Chadwick & Musa (2008)	NI	√	NI	NI	√	NI	NI	NI	NI	NI	NI	√	NI
Mumma (2004)	NI	√	NI	NI	NI	NI	Ni	X	X	NI	NI	√	NI
Mumma & Fluck (2016)	√	√	NI	√	NI	√	√	√	√	√	√	√	√
Mumma & Mooney (2007)	NI	√	NI	NI	NI	√	√	NI	√	NI	NI	√	√
Mumma & Smith (2001)	X	√	√	√	NI	√	√	X	√	√	X	√	X
Mumma, Marshall & Mauer (2018)	√	√	√	√	NI	√	√	NI	√	√	√	√	√
Muran, Samstag, Segal & Winston (1998)	N/A	√	√	NI	NI	√	√	NI	√	√	X	√	X

Authors (date)	Comprehensive	Relevant	Objective	Evidence-based	Material correctly interpreted	Clear relationships	Clear causal relationships	Explains provenance where necessary	Moderators clearly described	Coherent	Mechs of change clearly described	Provides basis for intervention	Provides rationale for prioritisation
Muran, Samstag, Ventur, Segal & Winston (2001)	N/A	√	√	NI	NI	√	X	NI	√	√	X	√	X
Muran & Segal (1992)	N/A	√	NI	NI	√	√	√	NI	√	√	X	√	X
Muran, Segal & Samstag (1994)	N/A	√	NI	√	NI	√	X	X	√	√	X	√	X
Nattrass, Kellett, Hardy & Ricketts (2015)	X	NI	NI	NI	NI	NI	NI	NI	NI	√	NI	NI	NI
Persons & Bertagnolli (1999)	NI	NI	NI	NI	NI	NI	NI	NI	√	NI	NI	√	NI
Persons, Bostrom & Bertagnolli (1999)	NI	NI	NI	NI	NI	NI	NI	NI	√	NI	NI	√	NI
Persons, Mooney & Padesky (1995)	NI	NI	NI	√	√	NI	NI	NI	NI	NI	NI	√	NI

Authors (date)	Comprehensive	Relevant	Objective	Evidence-based	Material correctly interpreted	Clear relationships	Clear causal relationships	Explains provenance where necessary	Moderators clearly described	Coherent	Mechs of change clearly described	Provides basis for intervention	Provides rationale for prioritisation
Persons, Roberts, Zalecki & Brechwald (2006)	NI	NI	NI	√	NI	√	√	√	NI	NI	√	√	√
Redhead, Johnstone & Nightingale (2015)	NI	√	NI	NI	NI	√	√	NI	NI	NI	NI	√	NI
Total	√ - 3 X – 7 NI – 16 N/A – 4	√ - 15 X – 0 NI - 15	√ - 4 X – 1 NI - 25	√ - 10 X – 2 NI - 18	√ - 6 X – 0 NI - 24	√ - 16 X – 0 NI - 14	√ - 13 X – 4 NI - 13	√ - 8 X – 6 NI - 16	√ - 10 X – 6 NI - 14	√ - 9.5 X – 1.5 NI - 19	√ - 6 X – 8 NI - 16	√ - 26 X – 0 NI - 4	√ - 5 X – 10 NI - 15

Note. *NI* No information; N/A Not applicable.

Evaluation of Study Quality

Quality assessment of studies included in this review was guided by the MMAT (Pluye & Hong, 2014), which was chosen as it permitted comparisons by score between studies of different designs, using criteria appropriate to each design. Overall, the mean score awarded to a study was 6.13 (range 2-7), with twenty-five studies (83.33%) achieving a score of six or seven, indicating that most included studies were of a high quality. Table 7 includes details of quality assessment for all studies, including the criteria for each study type and any potential difficulties identified.

In terms of MMAT criteria, the most prevalent source of potential bias, found in 19 out of 22 quantitative descriptive studies, was a lack of clarity over the representativeness of the sample used, which made it difficult to generalise findings beyond the study. The single mixed methods study included, Chadwick et al., 2003, was found to have methodological flaws, including a study design which did not allow the research questions to be answered as it assessed the effect of formulation but limited the time period of assessment to a brief period unlikely to be adequate for this purpose and affected by confounding factors. This meant that the study failed to meet standards for quantitative research, and in failing to account for these factors, did not interpret results appropriately. The integration of the qualitative and quantitative findings was also unclear. Two studies, both randomised controlled trials (Barris, 1996; Ghaderi, 2006) did not blind raters to condition, leading to a source of potential bias in interpretation of results.

In addition to the MMAT criteria, additional sources of bias were reviewed in order to obtain a more detailed picture of factors identified as relevant to this systematic review. Study and measure level bias were reviewed due to a high proportion of quantitative studies in the review. In terms of study level bias, a potential bias due to a clinician (the author) providing therapy to clients within the study and also reviewing their study outcome data was identified in three studies (Hess, 2000; Persons et al., 1999; Persons et al., 2006). Unintentional bias due to this lack of objectivity may have affected scoring and interpretation of outcome data. In terms of measure level bias, questions over the measures used to assess reliability were raised in two studies (Dudley et al., 2010; Persons

et al., 2006). In Persons et al. (2006) exclusively self-report measures were used to assess anxiety and depression, although they were used consistently pre- and post-therapy. While percentage agreement is a recognised method of assessing reliability (Eells et al., 1998) and so it has not been noted a source of bias the table below, it was noted that it is not as robust a measure of reliability as measures used in other included studies such as kappa, as it makes no adjustment to account for agreement due to chance. Percentage agreement was used in seven studies (Barris, 1996; Dudley et al., 2010; Ghaderi, 2006; Haarhoff et al, 2011; Kuyken et al., 2005; Persons & Bertagnolli, 1999; Persons et al., 1995). Questions over the validity of measures used to assess validity were raised in one study (Koerner, 1997). In this study formulation validity was assessed by asking a formulation expert to rate its clinical relevance to the patient. Notwithstanding their expertise, rating by a single person of a formulation's relevance is a limited measure of validity, and open to individual biases.

Publication bias can occur when the only studies included in systematic reviews are peer-reviewed papers or papers in the English language, which are features of the most commonly cited studies. Failing to include studies which do not meet these criteria can lead to relevant findings being excluded from systematic reviews which claim to present comprehensive assessments of a particular field. Unfortunately time limitations excluded the possibility of finding and translating any relevant studies not in English, although this review did include searches of grey literature and, as a result, included findings from three unpublished doctoral theses (Barris, 1996; Hess, 2000; Koerner, 1997). We also noted where studies reported non-significant findings, which indicates the absence of bias caused by reporting only significant findings.

In terms of sources of potential bias, in addition to those considered in MMAT criteria, the most dominant was the use of formulation models, measures or other assessment tools developed by the authors of papers which also analysed the results they produced (*n* = 20; 66.67% of all studies). This lack of independence had potential to introduce subtle biases which may have increased the apparent reliability of results or produced a failure to interpret them, or the framework within which they were interpreted, with sufficient critical distance.

Findings from this assessment of study quality were incorporated into the synthesis and discussion of review findings in order to give appropriate weight to studies according to their quality.

Table 7

Quality Assessment of Included Studies - Randomised Controlled Trials

Authors (date)	Clear research question (RQ)	Data allows RQ to be addressed	Randomisation appropriately performed	Groups comparable at baseline	Complete outcome data	Blind assessors	Treatment adherence	MMAT score	Study level bias	Measure level bias	Publication bias	Author using own model
Barris (1996)	Y	Y	Y	Y	Y	N	Y	6/7	N	N	GL, NS	√
Ghaderi (2006)	Y	Y	Y	Y	Y	N	Y	6/7	N	N	NS	
Kendjelic & Eells (2007)	Y	y	Y	Y	Y	Y	Y	7/7	N	N	NS	√

Note. *GL* Grey Literature; *NS* Reported non-significant results.

Non-Randomised Studies

Authors (date)	Clear research question (RQ)	Data allows RQ to be addressed	Participants representative of target population	Appropriate measurements	Complete outcome data	Confounders accounted for	Intervention correctly administered	MMAT score	Study level bias	Measure level bias	Publication bias	Author using own model
Persons, Roberts, Zalecki & Brechwald (2006)	Y	Y	CT	Y	Y	N	Y	5/7	OC	R?	NS	√

Note. *CT* Cannot tell; *OC* Author reporting own clients' outcomes; *R?* Unclear reliability of measures used; *NS* Reported non-significant results.

Quantitative Descriptive Studies

Authors (date)	Clear research question (RQ)	Data allows RQ to be addressed	Relevant sampling strategy	Representative sample	Appropriate measurements	Low risk of nonresponse bias	Appropriate statistical analysis	MMAT score	Study level bias	Measure level bias	Publication bias	Author using own model
Dudley, Park, James & Dodgson (2010)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	NS	N
Eells, Kendjelic & Lucas (1998)	Y	Y	Y	Y	Y	Y	Y	7/7	N	N	N	√

Authors (date)	Clear research question (RQ)	Data allows RQ to be addressed	Relevant sampling strategy	Representative sample	Appropriate measurements	Low risk of nonresponse bias	Appropriate statistical analysis	MMAT score	Study level bias	Measure level bias	Publication bias	Author using own model
Eells, Lombart, Kendjelic, Turner & Lucas (2005)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	N	√
Eells, Lombart, Salsman, Kendjelic, Schneiderman & Lucas (2011)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	N	√
Flitcroft, James, Freeston & Wood-Mitchell (2007)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	N	N
Haarhoff, Flett & Gibson (2011)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	N	√ - 1 measure out of 4
Hess (2000)	Y	Y	Y	N	CT	Y	N	4/7	OC	N	GL	N
Koerner (1997)	CT	N	Y	Y	CT	Y	Y	4/7	N	V?	GL, NS	
Kuyken, Fothergill, Musa & Chadwick (2005)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	NS	√ - quality ax
Mumma (2004)	Y	Y	Y	Y	Y	Y	Y	7/7	N	N	N	√

Authors (date)	Clear research question (RQ)	Data allows RQ to be addressed	Relevant sampling strategy	Representative sample	Appropriate measurements	Low risk of nonresponse bias	Appropriate statistical analysis	MMAT score	Study level bias	Measure level bias	Publication bias	Author using own model
Mumma & Fluck (2016)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	NS	√
Mumma & Mooney (2007)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	NS	√
Mumma & Smith (2001)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	N	√
Mumma, Marshall & Mauer (2018)	N	CT	Y	CT	Y	Y	Y	4/7	N	N	N	√
Muran, Samstag, Segal & Winston (1998)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	NS	√
Muran, Samstag, Ventur, Segal & Winston (2001)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	N	√
Muran & Segal (1992)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	NS	√
Muran, Segal & Samstag (1994)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	NS	√

Authors (date)	Clear research question (RQ)	Data allows RQ to be addressed	Relevant sampling strategy	Representative sample	Appropriate measurements	Low risk of nonresponse bias	Appropriate statistical analysis	MMAT score	Study level bias	Measure level bias	Publication bias	Author using own model
Nattrass, Kellett, Hardy & Ricketts (2015)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	NS	N
Persons & Bertagnolli (1999)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	NS	√
Persons, Bostrom & Bertagnolli (1999)	Y	Y	Y	CT	Y	Y	Y	6/7	OC	N	NS	√
Persons, Mooney & Padesky (1995)	Y	Y	Y	CT	Y	Y	Y	6/7	N	N	NS	√

Note. *CT* Cannot tell; *NS* Reported non-significant results; *OC* Author reporting own clients' outcomes; *GL* Grey Literature; *V?* Unclear validity of measures used.

Qualitative Studies

Authors (date)	Clear research question (RQ)	Data allows RQ to be addressed	Qual approach appropriate	Data collection method adequate	Findings derived from data	Interpretation substantiated	Coherent	MMAT score	Study level bias	Measure level bias	Publication bias	Author using own model
Kahlon, Neal & Patterson (2014)	Y	Y	Y	Y	CT	Y	Y	6/7	N	N	N	N
Morberg Pain, Chadwick & Musa (2008)	Y	Y	Y	Y	Y	Y	Y	7/7	N	N	N	N
Redhead, Johnstone & Nightingale (2015)	Y	Y	Y	Y	Y	Y	Y	7/7	N	N	N	N

Note. CT Cannot tell.

Mixed Methods Studies

Authors (date)	Clear research question (RQ)	Data allows RQ to be addressed	Adequate rationale for mixed methods	Effective integration	Adequate interpretation of data	Quant/qual inconsistencies addressed	Quality criteria for quant/qual met	MMAT score	Study level bias	Measure level bias	Publication bias	Author using own model
Chadwick, Williams & Mackenzie (2003**)**	Y	N	Y	N	N	CT	N	2/7	N	V?	NS	

Note. *CT* Cannot tell; *V?* Unclear validity of measures used.

Discussion

This review aimed to systematically evaluate the degree to which individual case formulation has been established as reliable and valid in cognitive behavioural therapy outcome research.

Evaluation of Formulation Reliability

Definition and Operationalisation of Reliability Assessment

Definitions of reliability varied between included studies. Eighteen studies assessed reliability and, in each one, inter-rater reliability was the focus of this assessment. However, the question of who this agreement should be with was understood, and so operationalised, broadly in three different ways. Seven studies used either a benchmark formulation produced by an expert, or else expert ratings, as the standard by which other formulations were evaluated. Five studies used the CFCCM (Eells et al., 1998), which categorises the content of case formulation as well as rating formulation quality. Five studies compared the ratings of multiple raters of the clinical relevance of self- or interpersonal scenarios. Each of these approaches points to a different underlying concept of case formulation. The studies assessing formulations against expert views imply by that comparison that the expert view is 'correct' and reliability is a measure of the degree to which other formulations resemble the 'correct' expert view. This definition implies a hierarchical view of cognitive case formulation, with 'experts' the ultimate arbiter of what is correct and so reliable. It can be argued that this is in direct opposition to the 'collaborative empiricism' (Beck, 1979, p.9) which is considered a hallmark of cognitive therapy. In contrast, the studies which assessed inter-rater views of self- or interpersonal scenarios did not place greater weight on any one rater's opinion of the formulation, with raters including the client in several studies. This reflects a more egalitarian view of cognitive case formulation, which aims for consensus in formulation rather than the imposition of a 'correct' view. The aim of the CFCCM was described as providing an explicitly atheoretical tool to reliably and comprehensively categorise the information clinicians include in formulations as well as an assessment of their quality (Eells et al., 1998). Based on factors common

to different formulation methods, the CFCCM represents an attempt to draw together existing knowledge without espousing any particular theoretical orientation, in order to be broadly applicable. This pragmatic approach to reliability implies an acknowledgment of the way that seemingly neutral concepts such as reliability may be shaped by underlying assumptions about what cognitive case formulation should aim to do, demonstrated here by these differing definitions of reliability, with implications for how they determine whether formulations are reliable. Overall, the conclusion drawn from this examination of definitions of reliability is that this process was not neutral, and that differences in the level of reliability obtained and the methods used to try and improve this were likely to reflect differences in the conceptualisation of case formulation as much as differences between formulations.

Common Factors in Reliable Formulations

This review confirmed previous research (Flinn et al., 2014) in finding that good inter-rater reliability could be obtained, and identifying the conditions which appeared to contribute to higher reliability, irrespective of the concept of reliability used. First, high inter-rater reliability was found the use of percentage agreement without correction for chance (e.g. Ghaderi, 2006) may have artificially inflated reliability scores in some studies. It was notably harder to obtain high reliability at a finer level of distinction, for example, in agreement on the specific cognitive mechanism maintaining a client's difficulties (e.g. Barris, 1996; 20% agreement; $\kappa = 0.05$).

Second, this review confirmed that good inter-rater reliability could be found for pooled groups of raters, but dropped significantly for single raters (e.g. Mumma & Smith, 2001; ICC (2,10) range = 0.832-0.945, ICC (2, 1) range = 0.330-0.633). A previous systematic review of formulation reliability (Easden & Kazantzis, 2018) noted that averaging or using the intraclass coefficient over multiple raters, which assumes that the sample is representative of the population, is likely to inflate reliability estimates. Our examination of study quality identified one of the most common limitations of studies in this review to be lack of clarity over the representativeness of the convenience samples

used, meaning that reliability estimates for pooled groups of raters in our included studies may also have been inflated.

Third, several studies in this review produced evidence in line with earlier studies (Flinn et al., 2014), for the greater reliability of descriptive over inferential content (Dudley et al., 2010; Eells et al., 1998), although this difference was less marked in some cases (Eells et al., 2005; Kuyken et al., 2005). Given that inferential aspects of formulation are internal cognitive processes, hence must be hypothesized based on available evidence (Persons, 1989), it may be argued that they are always likely to be less reliable than directly observable elements, rendering their apparent lower level of reliability unsurprising. The unpredictability of this process may also be increased by clinicians' unconscious use of heuristics such as confirmatory biases, where they look only for information to verify preconceived ideas, recency effects, or framing biases, in order to manage decision-making in the face of complex and voluminous client information (Kahneman, 2003; Kuhberger, 1998; Spengler & Strohmer, 1994). An alternative argument (Persons et al., 1995) is that cognitive theory itself may be the reason for this difference, as it is less elaborated in terms of the development of dysfunctional schema that drive different presentations than a comparative theory such as psychodynamic theory, for which greater inter-clinician reliability has been found (Crits-Christoph, Cooper & Luborsky, 1988). Formulation can only reflect the framework on which it is built, and if the relationships between presentation and likely inferential aspects are ill-defined, it follows that these aspects are less reliably formulated. Researchers have suggested that the level of definition may vary between psychological presentations, with the processes underlying depression better understood than those of anxiety (Bieling & Kuyken, 2003), although the studies in this review which demonstrated this effect included clients with difficulties including psychosis, anxiety and depression.

Increasing Case Formulation Reliability

Previous studies have identified factors which may increase formulation reliability, including experience, training and providing additional structure to guide formulations. Bieling and Kuyken

(2003) found only inconclusive results for the effect of experience on formulation reliability whereas in our review Kuyken et al. (2005) and Dudley et al. (2010) provided some support for a connection. Caveats with our findings were that findings in Kuyken (2005), based on a more than ten percentage points difference across experience levels, followed 'visual inspection of the percentage agreement' (Kuyken et al., 2005, p. 1197) rather using statistical methods; the finding of greater reliability among qualified clinicians in this study was not true for all formulation elements, for example, only one of the core beliefs identified was rated more reliably by experienced clinicians. Furthermore, both studies defined reliability as agreement with an expert formulation, which raises the possibility that, in this context, greater reliability simply represents experienced therapists' greater familiarity with expert views on 'correct' formulation. In other respects, this review did not fully support the findings of previous research. Earlier studies suggested training may improve reliability (Bieling & Kuyken, 2003), but our findings were more equivocal, with one study finding an association between possession of a PhD qualification and increased inter-rater reliability of formulation (Persons & Bertagnolli, 1999) and another finding no such association (Dudley et al., 2010). This review also produced only limited evidence for the idea considered in Easden & Kazantzis (2018), that more structured information may increase formulation reliability: additional structure did not increase reliability in Persons and Bertagnolli (1999), although the four studies conducted by Muran and colleagues (Muran & Segal, 1992; Muran et al., 1994; Muran et al., 1998; Muran et al., 2001), which used highly structured scenarios, produced high inter-rater reliability, albeit averaged across three raters.

Evaluation of Formulation Validity

A broad definition of validity in relation to case formulation may be to describe it as the degree to which case formulation captures and explains the factors contributing to and maintaining psychological distress. In psychometric terms, validity encompasses multiple specific meanings, including construct, discriminant and predictive validity. The evidence from studies within this review was that operationalising validity assessment within cognitive case formulation was

challenging, with factors such as outcome of formulation-driven treatment imperfect proxies for validity, in the apparent absence of more direct tools of measurement. Apart from the group of studies conducted by Mumma and colleagues which considered content, construct and/or predictive validity, studies were noted to maintain a rather narrow focus on construct validity, out all the subtypes of validity which could have been evaluated.

Formulation Validity and Treatment Outcomes

Nine studies examined the outcome of formulation-driven treatment. Considering these studies as relevant to validity assessment is based on the logic that a valid formulation may be associated with better outcomes; a limited and indirect assessment of formulation validity. While three studies based on naturalistic treatment within clinical settings found that cognitive formulation-driven treatment was associated with statistically and clinically significant improvements (Nattrass et al., 2015; Persons et al., 1999; Persons et al., 2006), control groups for these studies were either absent, non-clinical or also received formulation-driven treatment within a different setting. As they were naturalistic studies, potential confounding variables such as the use of additional therapies alongside CBT for some clients were also not excluded. As a result, the lack of clear operationalisation of validity in these studies meant that the good outcomes of these studies could not clearly be attributed to more valid formulation. Similar issues were identified with a randomised controlled trial considered in this review (Ghaderi, 2006), where the inclusion of individualised elements within the manual-driven treatment meant that there was insufficient distinction between formulation-driven and standardised treatment conditions. The potential impact of additional confounding variables such as the effect of therapist alliance meant that, again, differences between the outcomes of formulation-driven and standardised treatments could not be clearly attributed to formulation. Although the use of naturalistic treatment meant these studies had high ecological validity, the difficulty of using indirect measures of formulation validity was compounded by further poor operationalisation to the extent that drawing any firm conclusions

from these studies in respect of a relationship between outcome and formulation validity became problematic.

The remaining studies which considered outcomes of formulation-based treatment were similarly restricted to providing tentative support for formulation validity, again due to the use of indirect methods of assessment.. Muran et al. (1998) and Muran et al. (2001) found time-series data showed improvements for two parameters of their self-scenario formulations in line with nomothetic outcome measures, suggesting that they captured valid aspects of formulation; promising results, but constrained by being an association, rather than showing a more direct relationship. Evidence for a relationship between formulation-based treatment and better outcomes has been described as limited (Bieling & Kuyken, 2003), although more recent research is described as having advanced this to "moderate, but not overwhelming" (Persons & Hong, 2016, p. 97). For our purposes, this moderate association between treatment and outcome is not direct evidence of formulation validity, and even when evidence of correlations between specific aspects of a formulation and outcome is presented (e.g. Muran & Segal, 1992), although more convincing, this method is still indirect and so at best supports rather than proves formulation validity.

Formulation Validity and Clinical Utility

An alternative approach to evaluating formulation validity, though still indirect, was assessment of its clinical utility. The findings of Flitcroft et al. (2007) in relation to clinical validity, corroborated our earlier findings in relation to reliability by indicating that differences in theoretical orientation within cognitive therapy may also have influenced this nominally atheoretical assessment. Therapist responses to being asked which aspects of cognitive behavioural formulation for depression they found most useful appeared to reflect differences along theoretical lines, with the authors commenting that this may contribute to research findings of low-inter reliability, as noted previously in this review.

Perspectives on clinical utility from qualitative research indicated that perceiving the formulation to be accurate increased the sense of being understood (Kahlon et al., 2014; Morberg

Pain et al., 2008; Redhead et al., 2015), and that this perception was enhanced by collaboration between therapist and client (Kahlon et al., 2014). In the light of previous research indicating that the patient's acceptance of the treatment rationale predicts treatment outcomes (Addis & Jacobson, 2000; Fennell & Teasdale, 1987), these findings suggest a role for collaboration as a moderator of outcomes. Given the assumption that a valid formulation is also associated with better outcomes, further research exploring connections between collaboration and formulation validity may be merited. However, while assessment of validity in terms of clinical utility provided insight into factors which may increase a sense of formulation being useful, this was again an indirect assessment rather than a direct test of the validity of a formulation.

Psychometric Approaches to Formulation Validity

Finally, one group of studies, in particular those by Mumma and colleagues, took a much more directly psychometric approach to evaluating formulation validity, combining techniques such as experience sampling using ecological momentary assessment with statistical analysis to produce a method of validity assessment which used an individual's personal data to validate clinician's formulations. For some studies, this approach was used to validate partial formulations in the form of individual cognitive schemas, comparing the relative validity of expert vs. novice clinicians (Mumma & Mooney, 2007), although more complete formulations using functional analytical techniques were also analysed (Mumma et al., 2018). This innovative approach produced impressive results in terms of validity assessment, producing models which appeared to reflect participant's difficulties well, reflected in high scores for criterion and discriminant validity of items, and also explaining relationships between triggers and distress. Scores explaining distress were not always extremely high (e.g. 55% of hypothesized concurrent relations between triggers and distress supported; Mumma et al., 2018) but were reasonable, especially when considered as assessments of hypotheses. These methods relied on extensive collaboration with clients, and so did not take the 'expert' view of formulation correctness. They were relatively atheoretical in the sense that they did not clearly reflect a pre-determined idea of what may contribute to distress, with items included

based on both client and clinician contributions. Several of these studies had not been considered in recent systematic reviews of validity within cognitive behavioural therapy (e.g. Easden & Kazantzis, 2018), due to their small sample size. While these studies were often based on a single client, they were presented as tests of method, rather than suggesting the results of that particular study could be generalised. The method should also be tested by researchers independently of Mumma and colleagues to ensure bias due to testing a self-developed approach had not affected study interpretation. In practical terms, the time-consuming nature of this approach and specific method means that it may not always be easily applied within clinical practice, although the authors suggested ways to facilitate this (Mumma et al., 2018). Nonetheless, these studies represent an approach that was the most direct test of validity of all the studies included in this review, with promising results.

Evaluation of Formulation Quality

The inclusion within this review of studies assessing formulation quality within our evaluation of validity assessment was driven in part by the lack of distinction noted between these concepts in the research literature. For example, the description in Kuyken et al. (2005; p.1188) of quality assessment as verifying whether "the key constructs in case formulations meaningfully related to the person's presenting problems" might equally describe validity. By reviewing these studies, we hoped to unpick an apparent overlap and ambiguity in definitions of quality and validity not previously considered in systematic reviews, as well as to evaluate the extent of any separate contribution to formulation utility and therapy outcomes, with the overall aim of identifying more ecologically valid approaches to case formulation evaluation. Formulation quality was also considered to be a potentially more objective assessment of formulation than, for example, some assessments of reliability which deferred to a single expert view of the 'correct' formulation (Kuyken et al., 2005; Persons et al., 1995). Previous studies have established a potential association between higher quality formulation and outcome (Crits-Christoph, Cooper & Luborsky, 1988; Gower & Kuyken, 2011).

Measurement of quality within these studies was largely driven by use of quality assessment tools such as the CFCCM (Eells et al., 1998, used in five studies), the QRS (Kuyken et al., 2005, two studies), the CBT CC (Haarhoff et al., 2011, one study), as well as benchmark formulations (three studies). The CFCCM and CBT CC included assessment of the inclusion of key categories of content within the formulation, for example, descriptive and inferential information, plus the hypothesized underlying mechanism. The QRS focused on quality in terms of integration of information into a meaningful account of problems. The CFCCM included both content and quality ratings, with the eight specific quality ratings evaluating not only inclusion of material but qualitative aspects such as cohesion, how systematic the approach had been and fit between formulation and treatment plan. Some dimensions of CFCCM quality assessment, for example, degree to which the precision of language indicated that the formulation was tailored to an individual, suggested an overlap with assessment of validity. However, this was not true of every dimension, for example, coherence, which was an intrinsic aspect of the formulation, rather than a reflection of its relevance to a client. None of the studies presented evidence for their quality criteria, other than citing that it had been adapted from Strupp (Eells et al., 1998). However, it was noted in Eells et al. (1998) that the choice of criteria was driven by a desire to be objective and atheoretical, in contrast to the requirement to match the formulation of an expert noted in studies examining reliability (e.g. Kuyken et al., 2005). Findings from the majority of studies were that quality was poor to average (e.g. Eells et al., 1998; Haarhoff et al., 2011; Kuyken et al., 2005), which may be connected to the fact that studies in this category consistently used limited source information, such as written vignettes from which clinicians had to derive formulations, and so had low ecological validity. The conclusion drawn was that quality and validity appeared to have been defined and evaluated as distinct aspects of formulation for many dimensions, but that there was a degree of overlap found.

Our development of an independent quality ratings tool in order to evaluate formulation quality by standardised criteria across all included studies was significantly limited by the lack of

formulation detail within those studies. As a result, it was not possible to obtain further insight into the understanding of quality and its relationship to reliability and validity from this assessment.

Limitations

This systematic review is likely to have been limited by the definition of cognitive and behavioural therapy used. While it was necessary to use a clear definition for the search, and our definition was transparently stated, it was derived from a specific text book, the Handbook of Cognitive-Behavioural Therapies (Dobson & Dozois, 2019). Despite the aim to use the broadest possible definition, the definition used in this handbook and therefore in this review does not include approaches such as compassion-focused therapy, mindfulness-based cognitive therapy or meta-cognitive therapy. As a result, some interesting and relevant papers grounded in these approaches to formulation are likely to have been excluded .

This review did not distinguish between formulations developed for different problem types such as depression, anxiety and obsessive-compulsive disorder. As a result, potential differences in the reliability and validity of formulations developed for these heterogenous problem types may have been obscured. As noted earlier in respect of differences in reliability between descriptive and inferential aspects of formulation, where cognitive theory itself is less developed, a reduction in the reliability of formulations which rely on that theory is likely to follow. By the same rule, differences between the degree of development and elaboration of cognitive models for different problems types may have led to similar differences in the reliability and validity of their formulations. However, these potential differences were not identified in this review and so represent a possible area of refinement in future reviews of formulation reliability and validity.

Further limitations include the small samples used by many included studies, which confined the possibility of drawing statistical inferences. As identified in our quality assessment, the representativeness of samples was unclear in most studies, which made generalising from results difficult, albeit a high percentage of studies did not aim to generalise but rather test potential methods of reliability and validity assessment. The often limited and potentially ecologically invalid

source material, for example, written vignettes, along with constraints on the context of formulation, for example being asked to provide a 'think-aloud' formulation in five minutes, is likely to have hindered therapists to formulate in the way they might do in clinical practice. Understanding of the ways in which formulation may operate was sometimes poorly reflected in studies (e.g. Chadwick et al., 2003, which limited measurement of formulation impact to the session in which it was received), and so is likely to not have captured the information it claimed to review. A significant number of studies were tests of methods or models created by the authors, leading to the possibility of unintentionally biased interpretations of results.

Conclusions

This systematic review of the reliability and validity of cognitive and/or behavioural case formulation confirmed several key findings of previous research, including the difficulty of achieving good levels of inter-rater reliability between formulations, the way that this was often achieved only in fairly discrete contexts or for parts of the formulation and the degree to which this depended on agreement with a specific definition of reliability, which fit poorly with the idea of a reliability as a neutral psychometric term. Assessment of validity uncovered similar challenges, especially in terms of the operational definition of validity employed, with indirect proxies for validity such as outcomes often used, limiting insight. A further observation is that an assessment of validity which is possible only at the end of treatment is limited in terms of its clinical utility. Given the narrow way in which validity was operationalised in most studies within this review, future research would benefit from seeking to broaden the operational definition of formulation validity beyond construct validity to other subtypes, being clearer in defining the type of validity under consideration and especially by operationalising validity in these assessments more precisely, directly and in ways which more clearly exclude the potential influence of confounding factors so that study findings can be more definitively linked to the validity of formulation. In this respect, Mumma and colleagues' work presented a promising approach to validity assessment during treatment based on a combination of collaborative data collection and psychometric analysis, and including evaluation of content,

construct and predictive validity, although clinical implementation of this method may be complicated due to resource constraints. Finally, in this review, the effect of problem type on formulation reliability and validity was not considered. However, it appears possible that, given the different models used to formulate different psychological presentations, as well as potential differences in the types of symptoms observed and the ease of making inferences from them, that differences in the reliability and validity of formulations depending on problem type may be found, and that evaluation of these potential differences should form a clear part of any future review.

Overall, our findings indicated, as noted in previous reviews (Easden & Kazantzis, 2018) that attempts to translate the psychometric concepts of reliability and validity and apply them within the context of case formulation are problematic. Case formulation is not a psychometric concept and assessment in these terms provides an inevitably limited view of how well it achieves its aims. As noted by Persons and Tompkins (1997, p. 311), “the goal is not to find the “correct” formulation but to become skilled at generating hypotheses and using them to formulate intervention strategies". Reflecting on their study, Kuyken (2005) noted that a formulation being reliable with others, does not mean that it is either valid or of high quality. Our conclusion goes further: When different clinicians generate different formulations by focusing on different aspects of a client’s presentation, but both formulations may be considered to be valid, these studies point to the breakdown within this context of the psychometric assumption that reliability is a necessary component of validity, illustrating the insufficiency of an assessment of case formulation based on standards designed for psychometric measures.

However, it remains the case that if case formulation aspires to scientific rigour, it must demonstrate that it is more than unjustified interpretation. A possible area of future research indicated by this review is in the definition and assessment of ‘quality’ within case formulation. Research has demonstrated connections between high-quality case conceptualisations and better client outcomes (Abel et al., 2016; Crits-Christoph et al., 1988). Another important attribute of quality assessment is its aspiration to be objective in identifying the attributes of a ‘good’

formulation, in contrast to some of the reliability assessments reviewed (e.g. Kuyken et al., 2005; Persons et al., 1995). Although our own quality assessment of formulations was unable to produce insight into this area, tools used in included studies such as the CFCCM demonstrated attributes, such as being neutral or at least transparent in terms of approach, which this review has indicated are important in assessing objectivity. A recent review of measures assessing the quality of case conceptualisation identified the range of definitions of quality used within existing tools (Bucci et al., 2016). As recommended in that study, further research is required to clarify the processes and mechanisms by which formulation quality contributes to outcomes, and so determine the most appropriate dimensions by which to assess it.

Mapping Individual Symptom Networks for Post-Traumatic Stress Disorder

Abstract

The network approach to psychopathology is a way of understanding psychological distress which conceptualises mental health disorders as arising out of causal interactions between symptoms, rather than due to an underlying cause. Within this approach, mental health difficulties are represented by networks of symptoms which co-occur within a short time (contemporaneous networks) or predict one another's occurrence (directed networks). Networks can be constructed through gathering relevant data from individuals during their daily lives and using that data to represent their personalised networks of symptoms.

In this study we used this approach to gather symptom data from individuals diagnosed with post-traumatic stress disorder (PTSD) and construct their contemporaneous and directed networks in order to address two questions: whether these network-derived formulations resembled clinicians' cognitive formulations for the same individuals in terms of complexity and relationships between elements of the problem, and whether the networks could demonstrate a lead-lag relationship between cognitive and affective variables, in accordance with the cognitive theory of PTSD (Ehlers & Clark, 2000).

The findings were that clinician-derived formulations were more complex than network-derived formulations and provided clearer explanations for psychological distress, but that the relationships depicted by network-derived formulations were exclusively idiographic and founded on participant's personal data, whereas relationships in the clinicians' formulations were not clearly as personalised and evidence-based for each individual. Nonetheless, these findings supported the potential for network-derived formulations to provide insight complementary to clinicians' formulations within clinical settings.

We found too few cognitive-affective relationships in the directed networks estimated to support our claim of a lead-lag relationship between these variables. However, our study functioned as a proof-of-principle that using network analysis to operationalise theoretical constructs was justifiable. Future research using more personalised variables was recommended.

The network approach to psychopathology is a novel way of understanding mental distress, which has gathered significant research interest as an alternative to medical models of psychological difficulties (Fried et al., 2017). First described just over a decade ago (Borsboom, 2008), it proposes that mental health disorders arise as a result of causal interactions between symptoms, rather than because of an underlying cause (Greene et al., 2018); put simply, that symptoms constitute rather than reflect mental disorders (McNally et al., 2015). Within this model, symptoms are defined more broadly than within medicine, encompassing biological, social and psychological factors (Borsboom, 2017).

In practice, clinical psychologists work toward representing relationships between variables which interact to cause mental health difficulties by developing clinical formulations. Formulation is the means through which psychologists apply their knowledge of psychological theory and research to the particular difficulties experienced by an individual, in order to determine appropriate intervention. It is a core part of clinical practice, with proficiency in formulation a required skill for practitioner psychologists (HCPC, 2015). As research into potential applications of the network approach proceeds, piloting of network analysis within clinical settings has also started, with some positive results (Kroeze et al., 2017). If network analysis is to be developed as a clinical tool, better understanding of its distinctive qualities and role in relation to clinical formulation would be helpful in order to shape its use appropriately within this context. A vital question is whether an empirical process like network analysis can capture the same complex range of information contained within a formulation. The study of PTSD is a promising area within which to test out these ideas as it contains well worked out general models which would serve as a common basis from which to compare both network and therapist-derived formulations.

The particular capacity of network analysis to model interactions between specific symptoms has created another opportunity, which is to clarify aspects of psychological theory which have remained hypothetical until now due to the lack of a methodology sufficient precise to explore

them. Within PTSD one such opportunity to obtain corroborating evidence for a theoretical mechanism lies at the centre of the widely referenced cognitive model (Ehlers & Clark, 2000). Central to this model's explanation of what maintains PTSD symptoms is the hypothesis that dysfunctional appraisals of the traumatic event and/or its sequelae act to create an overwhelming emotional reaction, specifically a current sense of threat. Some aspects of this model have been confirmed through research, for example the sense of current threat (Regambal & Alden, 2012). However, understanding of the central relationship between cognitions about the event and the emotional response could benefit from the insight into inter-variable dynamics offered by network analysis. The network approach could facilitate this level of analysis through modelling contemporaneous and temporal networks. Contemporaneous networks depict the degree to which components of a symptom network are partially correlated within the same time interval, in other words the degree to which they co-occur, indicating a connection between those variables. Temporal networks model variable dynamics over time, by allowing the degree to which a variable is predicted by another at the previous time point to be calculated.

In the last few years, studies using analysis of these networks have made progress in clarifying relationships between symptoms within PTSD. In Greene et al. (2018), both longitudinal and cross-sectional data was gathered from a general population sample of Israeli citizens during a period in which they were experiencing rocket fire, and contemporaneous, temporal and cross-sectional models of their symptoms estimated using network analysis. The inclusion of longitudinal data allowed the temporal dynamics of symptoms of PTSD in this population to be analysed, including cognitive and affect variables. Findings included identification of a feedback loop in the temporal network, in which negative emotions predicated negative beliefs, which then predicted thought avoidance, which then predicted negative emotions. Their subsequent study in Greene et al. (2020) aimed to understand the relationship between negative emotions and the Diagnostic and Statistical Manual of Mental Disorders [DSM-5] (5th edition; APA, 2013) symptom clusters of PTSD. Although specific negative emotions such as sadness and despair were modelled, cognitions were

not included as a separate set of variables, but only as part of a symptom cluster of 'Negative Alterations in Cognitions and Mood' (NACM), which meant that this study could not offer further insight into the particular relationship between cognitions and emotions within either the contemporaneous or directed networks. The later study did find contemporaneous connections between sadness, loneliness, despair, anger and helplessness and NACM and that negative emotions were predicted by symptom clusters, particularly NACM and arousal.

Despite the useful insights provided by research to date, it is clear, therefore, that an opportunity remains to use the network approach to operationalise the cognitive theory of PTSD by carrying out a study enabling analysis of the relationship between cognitions and emotions for individuals with a diagnosis of PTSD through estimating their contemporaneous and directed symptom networks. This was the purpose of our study, which aimed to be a proof of principle study focused on this central interaction. As we did not wish to exclude potentially valuable data by predicting the direction of the relationships between cognitive and affect variables, we used a non-directional hypothesis. Further, in focusing on cognitions and emotions our study aimed to expand the scope of the type of variables modelled within network analytical studies in order to provide a fuller account of PTSD within this approach. While research into PTSD using network analysis has now been conducted in a wide range of participant groups, including survivors of terrorist attacks (Birkeland & Heir, 2017), refugees (Spiller et al., 2017) and adult survivors of childhood sexual abuse (McNally et al., 2017), variables analysed have been almost exclusively DSM-5-defined PTSD symptoms (Bryant et al., 2017) or symptom clusters (Doron-LaMarca et al., 2015). Ehlers and Clark's cognitive model (2000) includes a range of components such as appraisals, affect, and the nature of the traumatic memory; formulations draw together a rich tapestry of elements such as historical precipitating factors, contextual moderators and core beliefs as well as symptoms. Without demonstrating the capacity to include variables of this broader kind, it is difficult to see how network analysis can aspire to the same kind of explanatory power as these approaches.

We used ecological momentary assessment (EMA; Stone & Shiffman, 1994) for data collection in our study. This method has been used frequently in recent research using network analysis (Bak et al., 2016; Bringmann et al., 2013) and involves collection of data from participants at regular and frequent intervals, often using short surveys delivered to a mobile phone or tablet, in which participants are typically asked to provide information about the aspect of their life being researched 'in the moment'. For example, Kramer and colleagues (2014) provided participants in their study of psychotic symptoms and mood states with a wristwatch that beeped ten times a day for five days in order to prompt data collection. Wichers and colleagues (2014) collected data on mood state ten times daily in a study of depression. Advantages of EMA are that data is collected in the context of the participant's daily life, so has high ecological validity and is likely to be accurate, being less subject to recall bias (Santangelo et al., 2013). This is likely to be important for a project in which variables include highly labile components such as emotions. A potential disadvantage is that the repeated data collections can be wearisome for participants, resulting in missing data.

In conclusion, our study aimed to collect data using EMA from participants with a diagnosis of PTSD who had active symptoms and were in treatment during data collection. Two personalised symptom networks per participant were estimated from this data; a contemporaneous network to examine symptoms at one time point, and a temporal network to examine relationships between symptoms over time. These networks were evaluated, alongside a clinician-provided cognitive case formulation, in order to test two hypotheses:

Hypothesis 1: Network analysis-derived formulations will resemble therapist-derived formulations in terms of complexity and relationship between elements of the problem.

Hypothesis 2: There will be a 'lead-lag' relationship between changes in cognitive and affect variables.

Method

Participants

Sample

Our target for recruitment was 4-12 participants. This is a modest sample size, but was justified on the grounds that previous, larger studies using network analysis had not aimed specifically to operationalise a theory. As this was our aim, we needed to recruit a very specific, and also difficult-to-recruit sample to allow this analysis. Recruitment in excess of what was needed to prove the principle would waste participant and research effort, especially if the study did not produce the expected findings. For this reason, we decided to recruit only what was necessary to achieve the specific aim of this study.

Five people participated in the study. Demographic details are provided in Table 8. The sample was 100% female, 60% Caucasian and ranged in age from 21 to 51 years (*m* = 33.20, *SD* = 11.03). Two (40%) participants met criteria for at least one comorbid diagnosis, which included depression (*n* = 1), generalised anxiety disorder (*n* = 1) and borderline personality disorder (*n* = 1).

Table 8

Participant Demographics

Participant	Age	Gender	Ethnicity	Diagnoses
101	35	Female	White British	PTSD
102	21	Female	White British	PTSD, Depression, Generalised Anxiety Disorder
103	37	Female	White British	PTSD, Depression, Borderline Personality Disorder
104	51	Female	Indian	PTSD
105	22	Female	Middle Eastern: Arab	PTSD

Inclusion criteria were being adult (aged 18+) with a primary diagnosis of PTSD, at any stage of therapy for PTSD within a National Health Service [NHS] service, and having a clinician willing to

submit a cognitive formulation for them. Participants were recruited as a purposive sample from NHS specialist services providing treatment to people with a diagnosis of PTSD. In order to achieve our required sample it was necessary to prioritise participant eligibility and willingness to participate without limiting the sample to particular clinicians. As a result, each participant was treated by a different therapist, who had varying levels of experience, as documented in Table 9. Participating therapists had the same level of experience with the particular formulation model used in this study, as it was devised specifically for this project.

Table 9

Clinician Experience

Participant	Clinician – Years of Post-Qualification Experience
101	20
102	15
103	3.5
104	7.5
105	1

Recruitment

NHS ethical approval was obtained from South Central Berkshire B Research Ethics Committee of the NHS Health Research Authority as an amendment to a previous study supervised by Dr. Gary Brown (Appendix A). The study included four phases: a) recruitment of clinicians and clients, b) completion of formulation training and submission of client formulation and predicted symptom relationships table by clinician, completion of baseline measures and predicted symptom relationships table by client, c) 14-day EMA sampling period by client, and d) completion of end of study measures by client. In recognition of the intensive data collection required by this study, participants were compensated for their time, up to a limit of £6/day.

Following ethical approval and agreement of the NHS service to participate, clinicians approached clients fitting recruitment criteria to provide them with the Participant Information

Sheet (Appendix B). Potential participants were then contacted by the researcher who explained the study and took informed consent if they wished to proceed. All participants completed consent forms (Appendix C) and were free to withdraw at any point.

Measures

Three self-report measures used within NHS services providing treatment for PTSD were completed by participants in order to establish symptom presence and severity at baseline and assist in creating their personalised survey. These measures evaluated the common symptoms of PTSD as defined in DSM-5 (APA, 2013), as well as common cognitions relating to traumatic experiences and responses to traumatic intrusions; they were the PTSD Checklist for DSM-5 (PCL-5; Weathers et al, 2013; Appendix D), the Post-traumatic Cognitions Inventory (PTCI; Foa et al., 199*9*; Appendix E) and the Response to Intrusions Questionnaire (RIQ; Clohessy & Ehlers, 1999; Murray, Ehlers, & Mayou, 2002; Appendix F).

PTSD Checklist for DSM-5 (PCL-5; Weathers et al, 2013)

This 20-item self-report measure is used to assess the presence and severity of PTSD symptoms over the last month, as specified in the DSM-5 (APA, 2013). Items correspond with the criteria used in the DSM-5 to make a diagnosis of PTSD. In this study it was used as a baseline and end of study measure to verify symptom severity and indicate likelihood of meeting diagnostic criteria for PTSD at those stages. Items are rated on a five-point scale (0 = *Not at all* to 4 = *Extremely*). Items are summed to create a total severity score with the range 0-80; higher scores indicate greater symptom severity. Scores above 31-33 are considered indicative of PTSD, alongside assessment of specific symptom severity against diagnostic criteria, although scores on this screening tool are not recommended as a basis for diagnosis of PTSD without additional clinical assessment. The authors have documented good test-retest reliability (range 0.66-0.96), internal consistency (α range 0.83-0.98), convergent validity (correlations with other PTSD measures from 0.62-0.93) and discriminant validity (correlations with measures of related constructs below 0.87) (Blevins et al., 2015).

Post-Traumatic Cognitions Inventory (PTCI; Foa et al., 1999)

We used the short 33-item form of the PTCI, which is limited to the nonexperimental items included in scoring. This is a self-report measure used to assess posttraumatic cognitions, and it was used to evaluate strength of belief in cognitions in our study. It comprises three sub-scales: Negative Cognitions About the Self, Negative Cognitions About the World and Self-Blame. Items are rated on a 7-point Likert scale (1 = totally disagree, 7 = totally agree). The PTCI total score is the sum of the raw scores of the three subscales (range 33-231); higher scores indicate greater severity. Subscale scores are reported as means. Internal consistency for the three subscales has been reported as sound (Negative Cognitions about the Self, α - 0.97, Negative Cognitions about the World, α = 0.88, Self-Blame, α = 0.86) with test—retest reliability for a three-week interval reported as 0.80-0.86 for the three subscales, and the scale found to differentiate well between individuals with and without a diagnosis of PTSD (sensitivity = 0.78. specificity = 0.93; Foa et al., 1999). Internal consistency for the total scale has been rated as 0.96 (Wells et al., 2019).

Response to Intrusions Questionnaire (RIQ; Clohessy & Ehlers, 1999; Murray, Ehlers, & Mayou, 2002)

This 19-item self-report questionnaire is used to assess maladaptive responses to intrusions following a traumatic experience, specifically evaluating rumination, suppression and numbing which form subscales. It was used to assess affective responses in our study. Items are rated on a 4-point scale (0 = Never, 3 = Always) with authors recommending that mean scores for the total scale and subscales are reported. Scores increase with the strength of maladaptive responses. The RIQ has been shown to predict PTSD in several studies (e.g. Kleim et al., 2012; Wild et al., 2016). Internal consistency of the total scale (α = 0.82) and the subscales (rumination α = 0.91, suppression α = 0.93, numbing α = 0.86) were found to be good, although in a perinatal population not all of whom met diagnosis for PTSD (King et al., 2017).

Personal Symptom Survey

Each participant in the study provided data in response to a personalised survey during the data collection period. The format for each participant's survey was standardised and asked about four types of symptoms: core PTSD symptoms, cognitive symptoms, affective symptoms and contextual moderators. The first question checked for the occurrence of a trigger for symptoms since the last questionnaire, and asked the participant to describe it. Questions 2-6 were based on the PCL-5 and confirmed whether key symptoms associated with PTSD (memories, reexperiencing, distress/physical reactions, hypervigilance, feeling jumpy) had been observed since the last questionnaire. Responses were via a 5-point scale from 'Not at all' to 'Extremely'. Question 7 asked participants to rate their agreement with the three thoughts or beliefs with which they had indicated the strongest agreement on the PTCI, with possible responses ranging from 'Strongly disagree' to 'Strongly agree' on a five point scale. Question 8 asked them to rate the frequency with which they responded to intrusions using the three maladaptive strategies they identified as using most often on the RIQ. Possible responses ranged from 'Never' to 'Always' on a 4-point scale. If more than three items had top scores on either the PTCI or RIQ the items included were those the participant felt were most characteristic of their experience of PTSD and how they responded to symptoms. Questions 9 and 10 were left open for participants to indicate the presence/absence at the time of responding of personally chosen moderators which they felt affected their responses to symptoms. Appendix G provides an anonymised example of a personalised questionnaire. All personalised surveys consisted of 10-12 questions, which were designed to take 3-4 minutes to complete in total. Previous studies using network analysis to create contemporaneous and/or temporal models of mental health conditions have also included questions based on existing nomothetic measures of symptoms, as we did (David et al., 2018; Reeves & Fisher, 2000). The final version of each participant's personalised survey was agreed in a call between the participant and researcher prior to the start of EMA sampling.

Procedure

Participants

Socio-demographics. Basic socio-demographic information was collected including participants' age, gender, ethnicity and current mental health diagnoses.

Baseline Measures. Participants completed three standardised self-report measures, the PCL-5 (Weathers et al, 2013; Appendix D), the PTCI (Foa et al., 1999; Appendix E) and the RIQ (Clohessy & Ehlers, 1999; Murray, Ehlers & Mayou, 2002; Appendix F).

Predicted Relationships Table. Participants completed a table (Appendix H) which cross-tabulated the symptoms included in their personalised survey, using a cross to indicate any two different symptoms which they perceived to co-occur or be related in some way, in response to the question 'Please complete the table, below, by placing an X in any unshaded box where the two items in question are related to each other in your life'. Participants were free to indicate as many or few relationships between symptoms as they perceived existed.

14-day EMA Sampling Period. All participant surveys were hosted on the online platform Qualtrics (Version 2021), which is compliant with GDPR regulations. Following creation of their personalised survey, data collection proceeded using the EMA approach (Stone & Shiffman, 1994). Once the survey was agreed, participants started to receive their personalised survey via email to their mobile phone, five times a day for fourteen days, starting the day after their next therapy session. Surveys arrived every three hours within a 12-hour window. The start and end times for this window was decided by the participant, to best reflect their typical routine and experience of symptoms (e.g. from 9a.m. – 9p.m.). Questionnaires expired and could no longer be completed two hours after arrival to ensure spacing of responses. Surveys were time-stamped with the time of completion.

End of Study Measures. Following completion of the 14-day EMA sampling period, participants completed the same three standardised questionnaires completed at baseline (PCL-5, PTCI and RIQ) immediately after completing their last personalised survey.

Clinicians

Formulation. Clinicians whose clients had agreed to participate in the study completed an online training module in Individual Case Formulation (ICF; Hallam, 2013) for PTSD, which was the standard model for formulation used in the study. ICF is a formulation approach founded on principles of functional analysis, which aims to specify causal relationships, contextual moderators and mediators precisely. The ICF model for PTSD incorporates Ehlers and Clark's (2000) cognitive model of PTSD. As such it provided a common basis for comparison with network models, which sought to identify all the factors contributing to and maintaining dysfunction. Following training, each clinician uploaded an anonymised cognitive case formulation for their client following the ICF format.

Predicted Relationships Table. Clinicians also completed an identical table to their client (Appendix H), noting any relationships between symptoms included in that client's personalised survey that they perceived.

Network Data Analysis

Each participant self-reported their symptoms in response to the same personalised short questionnaire five times daily for a total of fourteen days (excepting the first participant, who requested to collect data for twenty-one days). This produced seventy data points per question, if there was no missing data. No standard procedure for power analysis in order to ascertain an appropriate sample size for network analysis has yet been devised, although a previous study has recommended 20-50 observations per subject in order to estimate edges in a directed network (Vrijen et al., 2018). Our sampling rate (five surveys per day) and the duration of the data collection period (14 days) were selected as likely to create sufficient observations to allow estimation of network models, allowing for missing data, balanced against the need to ensure that data collection was not too burdensome, which can also produce missing data. The protocol was piloted by a researcher ahead of the study start.

The interval between questionnaires, known as the window of measurement (Epskamp et al., 2018), was three hours. Analysis of data followed the method validated in Epskamp et al. (2018),

using vector autoregression (VAR; van der Krieke et al., 2015) to compute two personalised symptom networks per participant: a contemporaneous network illustrating whether symptoms predicted one another within the same window of measurement, and a temporal network illustrating whether symptoms predicted one another over time.

'Graphical VAR' (Wild et al., 2010) describes the modelling of contemporaneous effects of a VAR model using a partial correlation network. A graphical VAR package (https://cran.r-project.org/web/packages/graphicalVAR/graphicalVAR.pdf) was used to analyse our time series data and estimate the two networks required per participant by identifying the models best fitting the data. Both types of networks were depicted using Gaussian graphical models (Lauritzen, 1996), in which 'nodes' representing observed variables (symptoms) were connected by 'edges' representing estimated inter-symptom relationships. Edge weight corresponded to relationship strength, with thicker edges indicating stronger relationships. Regularization was achieved using LASSO with a gamma parameter set to .1. This is a comparatively low setting for a gamma parameter, which meant models would err in favour of finding edges, as we wanted to include as much potentially relevant data as possible, but which also increased the risk of false positives in terms of the edges found.

In contemporaneous networks edges depict the partial correlations between nodes still found after controlling for temporal effects and all other variables within the same window of measurement (Wild et al., 2010). The edges were not shown with arrowheads as contemporaneous networks are undirected. Edge colour indicated whether a correlation was positive or negative, with green edges positive, and red negative. Temporal networks were constructed using arrows to connect the preceding node to the following one, whenever one variable (symptom) predicted another in the next in the next window of measurement, which may also be described as a time-1 lagged association between symptoms. This temporal prediction, also referred to as Granger causality (Granger, 1969) is a necessary but not sufficient basis for causality; it does not prove causality alone but is observed in causal links as cause precedes effect. In both temporal and

contemporaneous network connections indicate potential causal relationships but are not proof of them. It is recommended that these networks are viewed as hypothesis-generating rather than confirmatory (Epskamp, 2018).

Interpretation of Results

Hypothesis 1: Comparison of Network and Therapist-Derived Formulations

Complexity. To compare the complexity of the network-analysis and therapist-derived formulations (represented in Figure 2), we used a definition of complexity taken from Eells et al., (1998) which states that complexity is "the degree to which the formulation takes into account several facets of the person's current problems and integrates these facets into a meaningful account" (Eells et al., 1998, p.148). This definition was chosen as it was a clear and specific description of complexity within formulation. It was operationalised in terms of three criteria, which were (1) the number of formulation elements included, (2) the types of information included and (3) the degree to which this information was integrated into a cohesive picture of the person's difficulties. To calculate the number of formulation elements included, we counted the number of nodes in network formulations and the number of discrete items of information in clinician-derived formulations, for example, 'Feeling trapped' in the formulation for participant 101. To analyse the types of information included, a list of specific observations provided by each formulation was compiled by the researcher, entered into a table, and the different information provided by each formulation approach compared. Analysis of the degree to which the information was integrated into a cohesive picture was based on researcher assessment of the extent to which each formulation included and linked information to provide an explanation of how symptoms arose and/or were maintained, without requiring additional interpretation. This assessment was also detailed in the table reporting the analysis of formulation complexity.

Relationships Between Elements of the Problem. We analysed the resemblance between network analysis-derived formulations and therapist-derived formulations in terms of relationships between elements of the problem by considering two areas i) the number of relationships included in the

formulation ii) the degree of similarity of the relationships identified between variables within the different formulation approaches. The number of relationships included in each formulation was determined by counting the number of edges (lines) in each network formulation and the number of lines linking formulation elements in each therapist-derived formulation. The degree of similarity of the relationships between variables was established on the basis of two criteria, which were i) the extent to which relationships were identified between the same variables within the two formulation approaches ii) the type of inter-variable relationship identified (e.g. causal, correlative).

Hypothesis 2: Relationship Between Changes in Cognitive and Affect Variables

In order to examine the direction of effect between changes in cognitive and affect variables within the network formulations, we first specified which variables within each formulation should be classified as the cognitive and affect variables of interest. Cognitive variables were defined as those variables selected for inclusion in the participant's personal symptom survey from the PTCI (Foa et al., 1999), identifiable within the network diagrams in this study as nodes with a label starting with '7'. Affect variables were those selected for inclusion from the RIQ (Clohessy & Ehlers, 1999; Murray, Ehlers, & Mayou, 2002), identifiable as nodes with a label starting with '8'. All relationships between a cognitive and an affect variable within participants' directed and contemporaneous networks, in either direction, were then listed for analysis.

Results

Standardised Measures

The purpose of the standardised measures was 1) to check participants met diagnosis for PTSD; 2) to provide a basis for individualised surveys; 3) to provide confirmation, context and explanation for any unusual results found in the longitudinal data if required. As pre and post measures were completed two weeks apart, and change in these scores was not a focus on the study, differences in pre-post scores were not calculated. All participants met diagnosis for PTSD at the start and end of data collection, according to their total and criteria scores on the PCL-5 (defined as a total score greater or equal to 33, scores of at least 2 on one criterion B and C item and two

criterion D and E items; U.S. Department of Veterans Affairs, 2021). Pre-post scores on the standardised measures are reported in Table 10.

Table 10

Pre-Post Scores on Standardised Measures

Participant	PCL-5		PTCI								RIQ							
			NCS *m*		NCW *m*		SB *m*		Total		Suppression *m*		Rumination *m*		Numbing *m*		Total *m*	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
101	60	59	4.48	5.00	6.86	6.14	6.60	5.60	175	176	1.83	2.00	1.38	1.63	1.40	1.80	1.53	1.79
102	72	74	5.95	6.10	6.86	6.71	4.40	5.80	195	204	2.50	2.50	1.50	2.00	2.20	2.00	2.00	2.16
103	78	73	6.24	5.81	6.86	6.43	6.60	5.80	212	196	1.67	2.17	1.38	1.88	2.20	2.20	1.68	2.05
104	48	54	6.00	4.71	6.86	6.29	5.20	4.00	200	163	1.33	2.00	1.88	1.75	1.40	1.40	1.58	1.74
105	66	47	5.38	4.48	7.00	6.71	6.40	5.80	194	94	2.67	2.17	2.38	1.50	2.20	1.60	2.42	1.74

Note. *PCL-5* PTSD Checklist for DSM-5 (Weathers et al., 2013); *PTCI* Post-Traumatic Cognitions Inventory (Foa et a., 1999); subscales for PTCI: *NCS* Negative Cognitions about Self; *NCW* Negative Cognitions about the World; *SB* Self-Blame; range for PTCI total score: 33-231; *RIQ* Response to Intrusions Questionnaire (Clohessy & Ehlers, 1999; Murray, Ehlers & Mayou, 2002).

Hypothesis 1

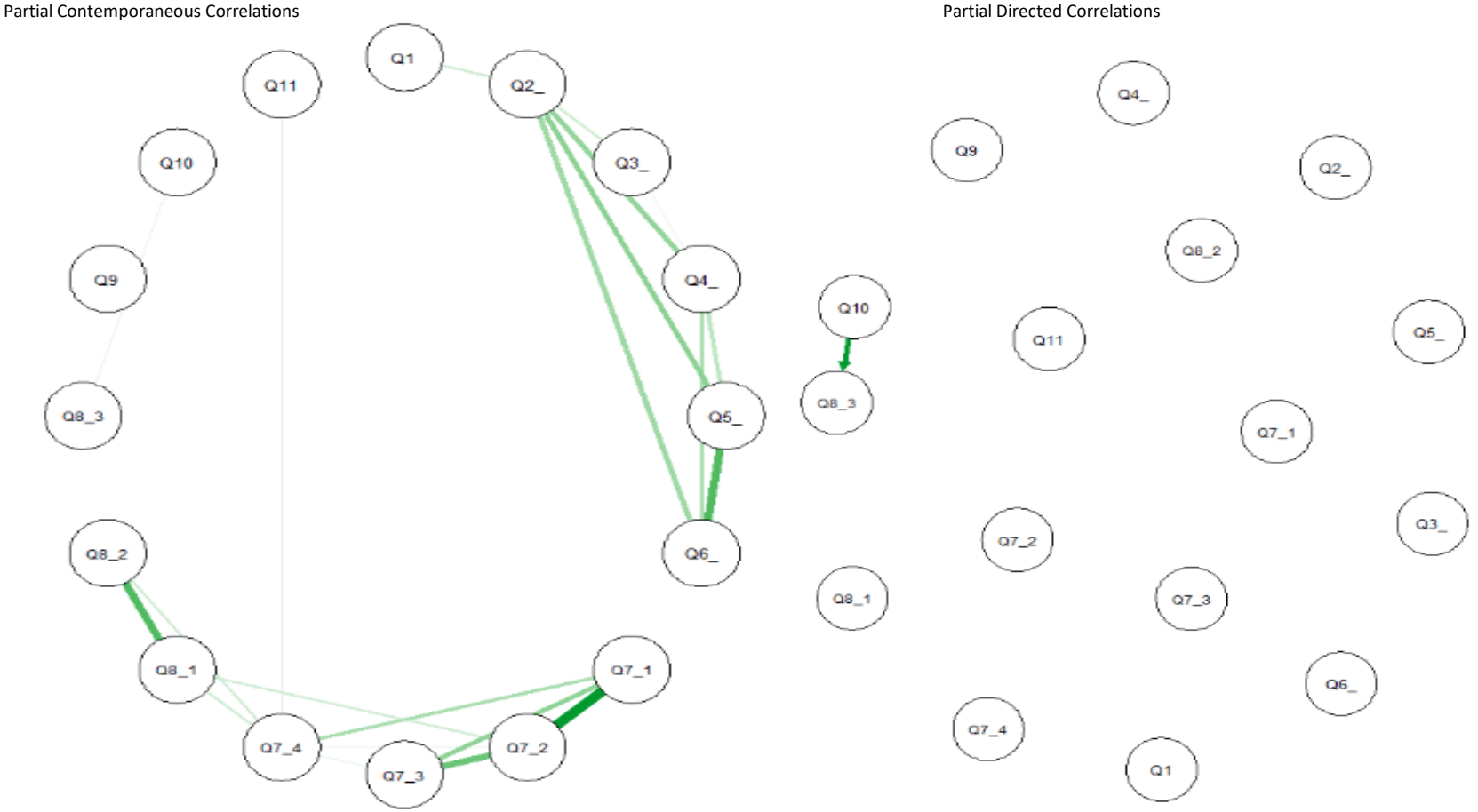
In hypothesis one, we proposed that network analysis-derived formulations would resemble therapist-derived formulations in terms of complexity and relationships between elements of the problem. The network maps and clinicians' formulations, which were based on Individual Case Formulation (Hallam, 2013), are represented in Figure 2. Clinicians' full formulation diagrams were anonymised using equivalent but fictional elements where material was deemed too identifying of participants.

Figure 2

Contemporaneous and Directed Networks and Clinicians' Formulations

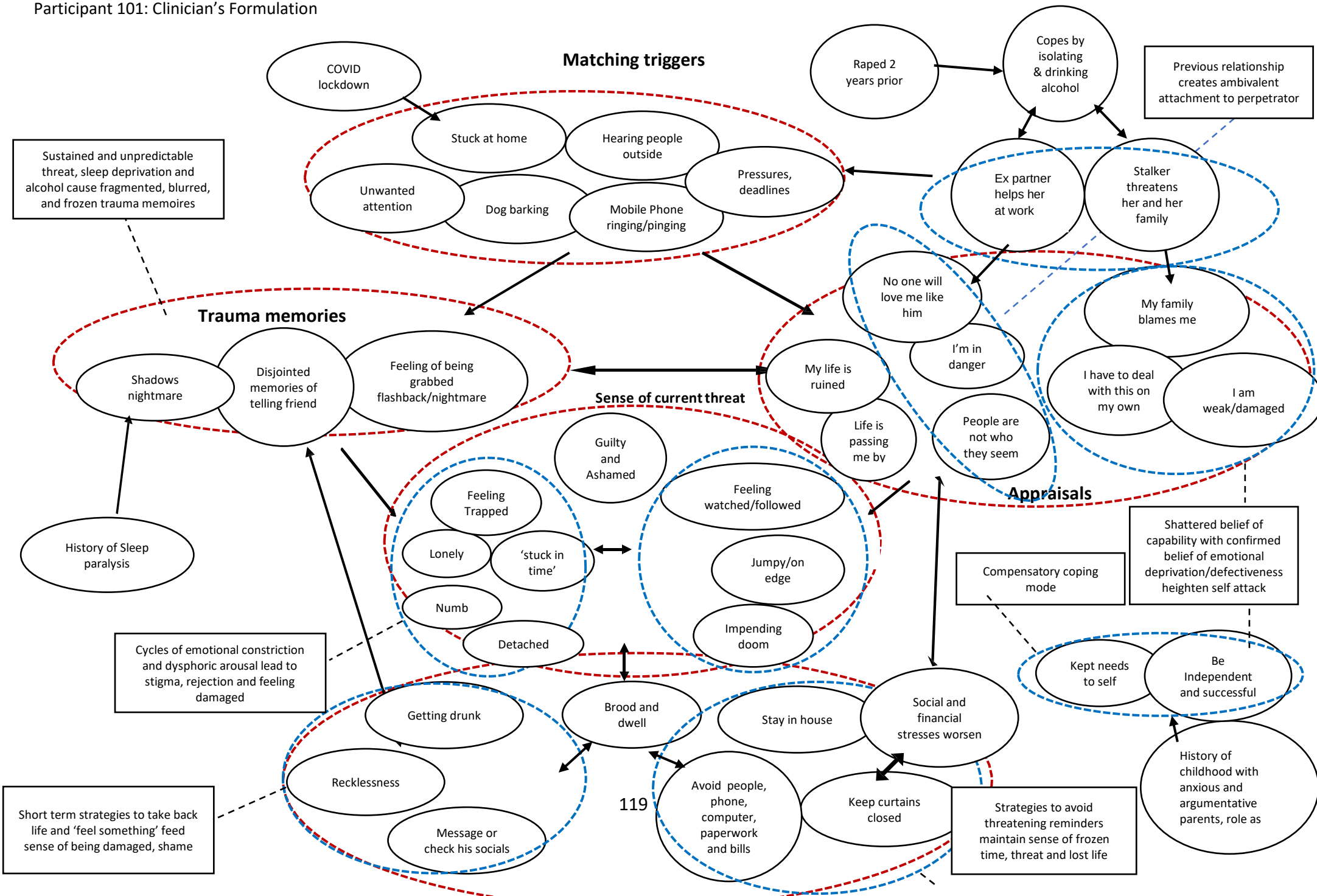
Participant 101: Contemporaneous and Directed Networks

Green = positive correlation Red = negative correlation



Q1 Trigger	Q7_1 Thought: People can't be trusted	Q8_3 Reaction: I drink alcohol, take medication or use drugs
Q2 Memory	Q7_2 Thought: You can never know who will harm you	Q9 I am tired
Q3 Flashback/Nightmare	Q7_3 Thought: The event happened to me because of the sort of person I am	Q10 I have had an alcoholic drink today
Q4 Upset / strong physical reaction	Q7_4 Thought: I feel isolated and set apart from others	Q11 I am with friends
Q5 Superalert	Q8_1 Reaction: I dwell on how I used to be before the event	
Q6 Jumpy / easily startled	Q8_2 Reaction: I dwell on what other people have done to me	

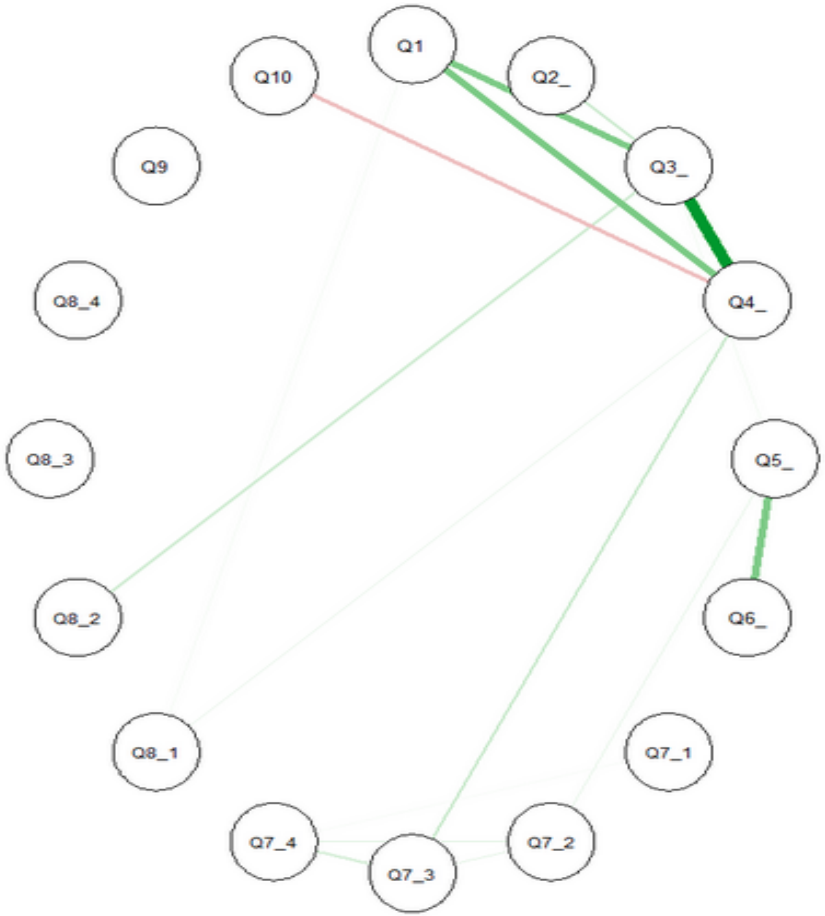
Participant 101: Clinician's Formulation



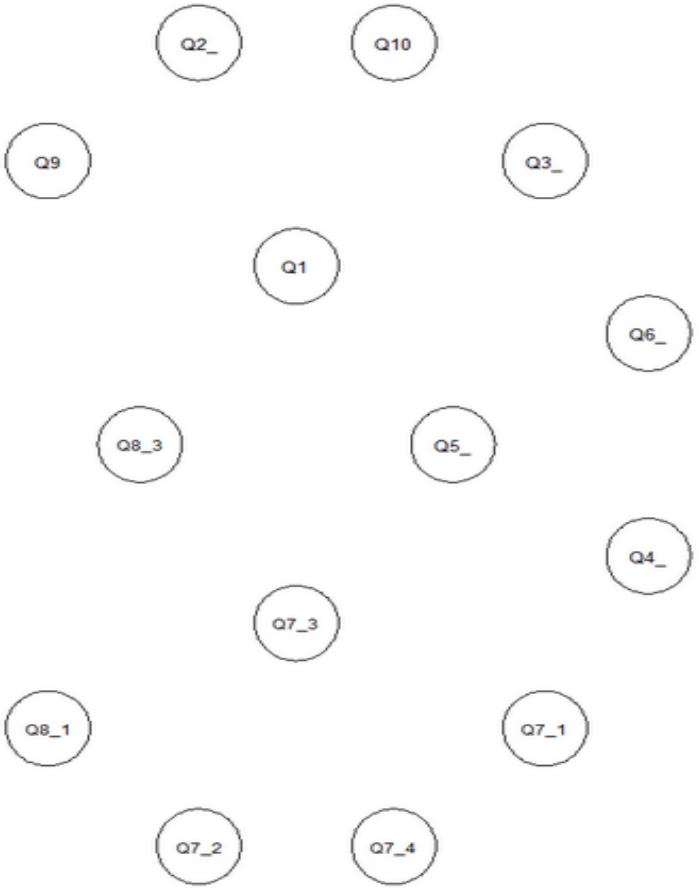
Participant 102: Contemporaneous and Directed Networks

Green = positive correlation Red = negative correlation

Partial Contemporaneous Correlations

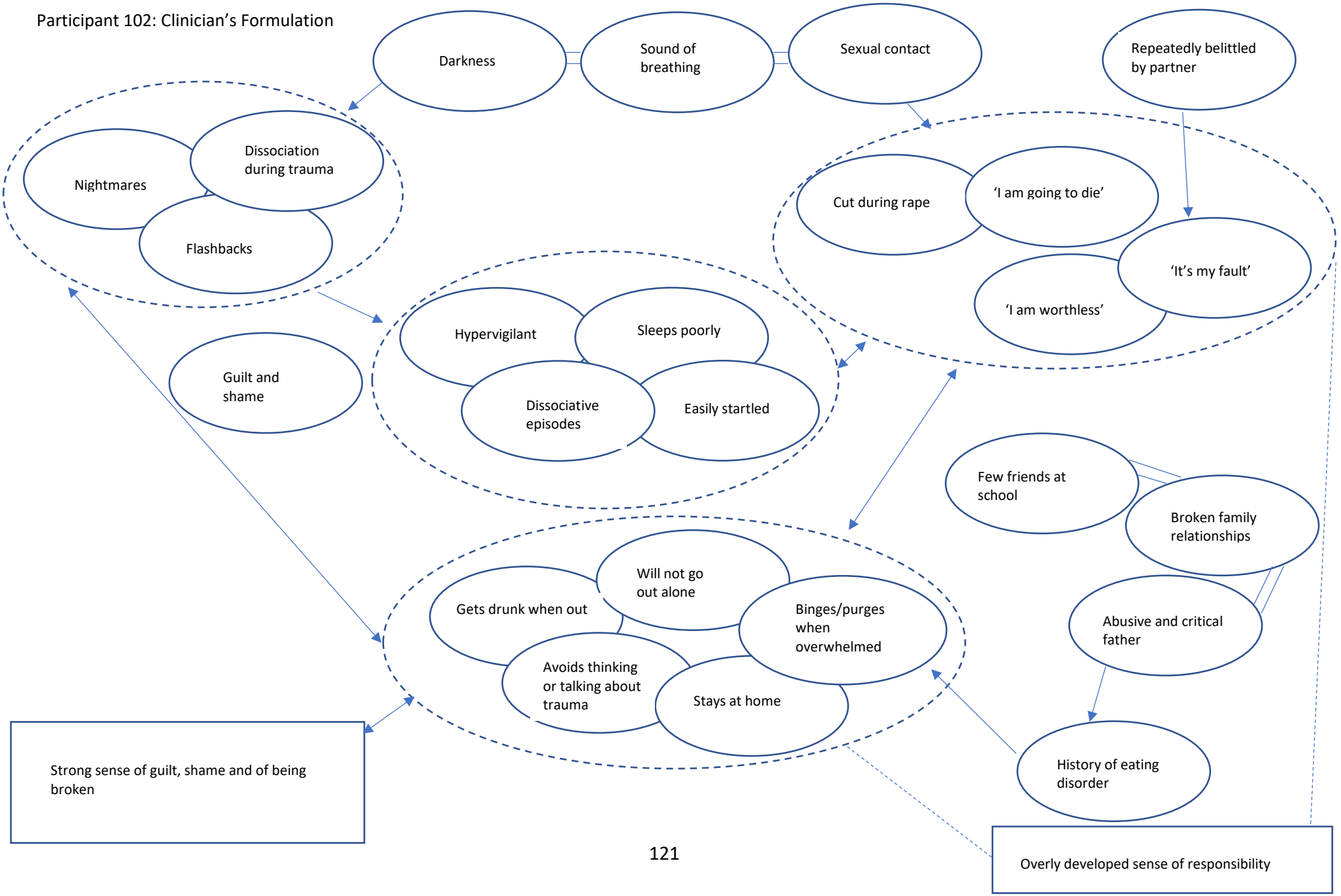


Partial Directed Correlations



Q1 Trigger	Q7_1 Thought: I am inadequate	Q8_3 Reaction: I try to push memories out of my mind
Q2 Memory	Q7_2 Thought: I have to be on guard all the time	Q8_4 Reaction: I work hard at keeping busy with other things
Q3 Flashback	Q7_3 Thought: You can never know who will harm you	Q9 I am alone
Q4 Upset / strong physical reactions	Q7_4 Thought: I feel like an object, not like a person	Q10 I have eaten a normal amount of food for my last meal
Q5 Superalert	Q8_1 Reaction: I dwell on what I should have done differently	
Q6 Jumpy / easily startled	Q8_2 Reaction: I numb my feelings	

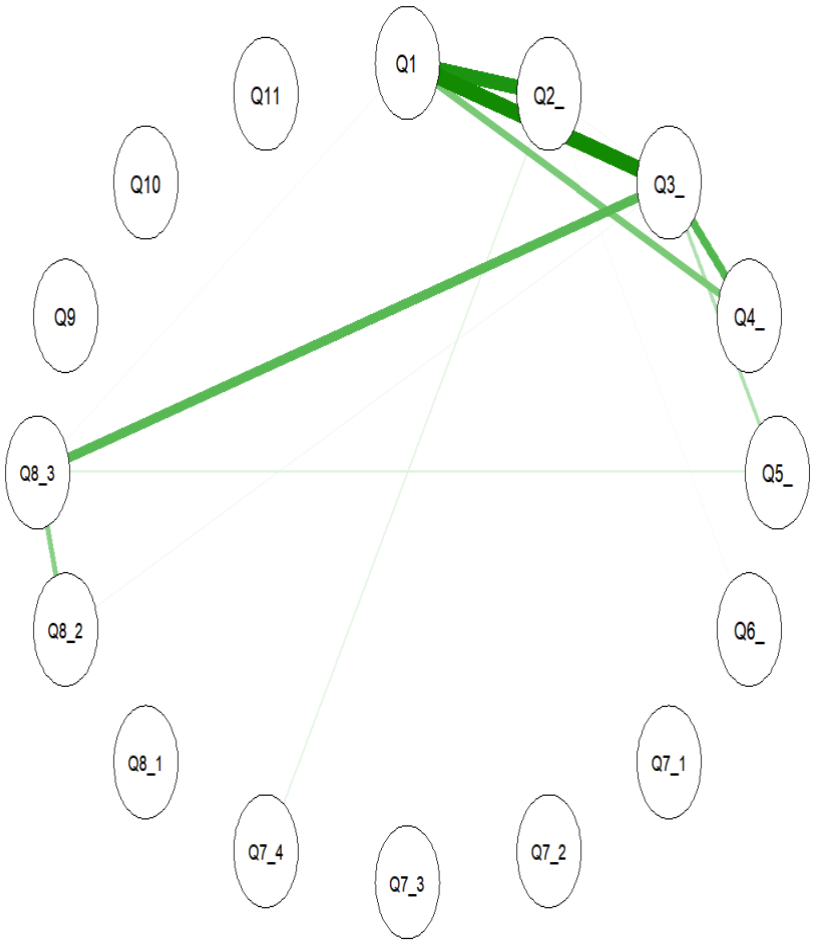
Participant 102: Clinician's Formulation



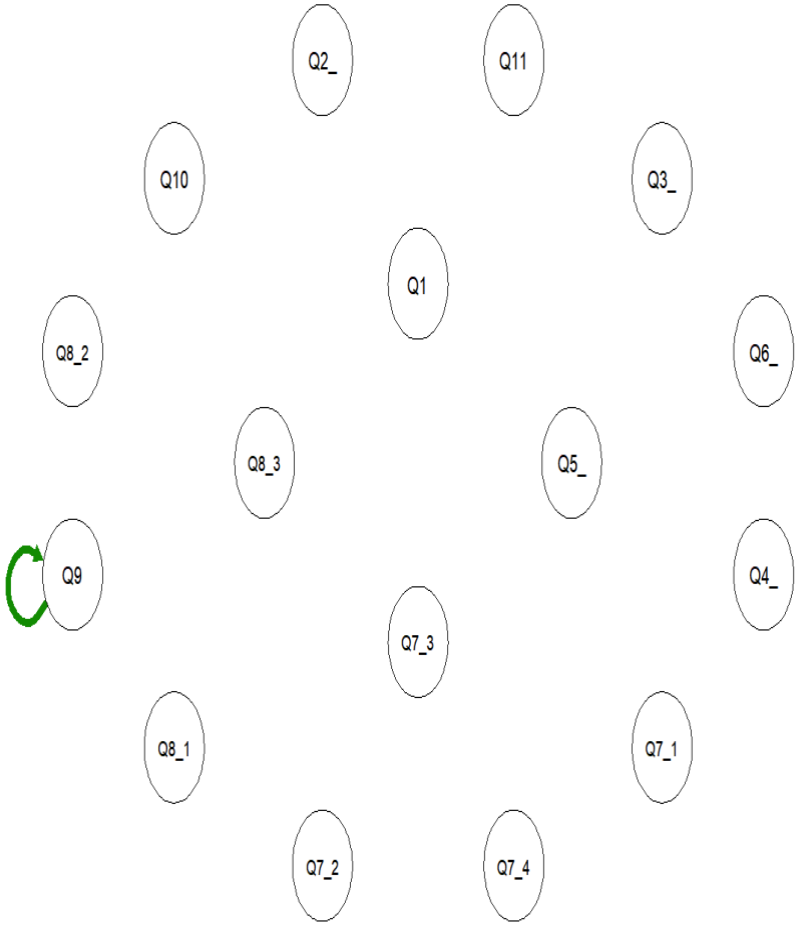
Participant 103: Contemporaneous and Directed Networks

Green = positive correlation Red = negative correlation

Partial Contemporaneous Correlations



Partial Directed Correlations



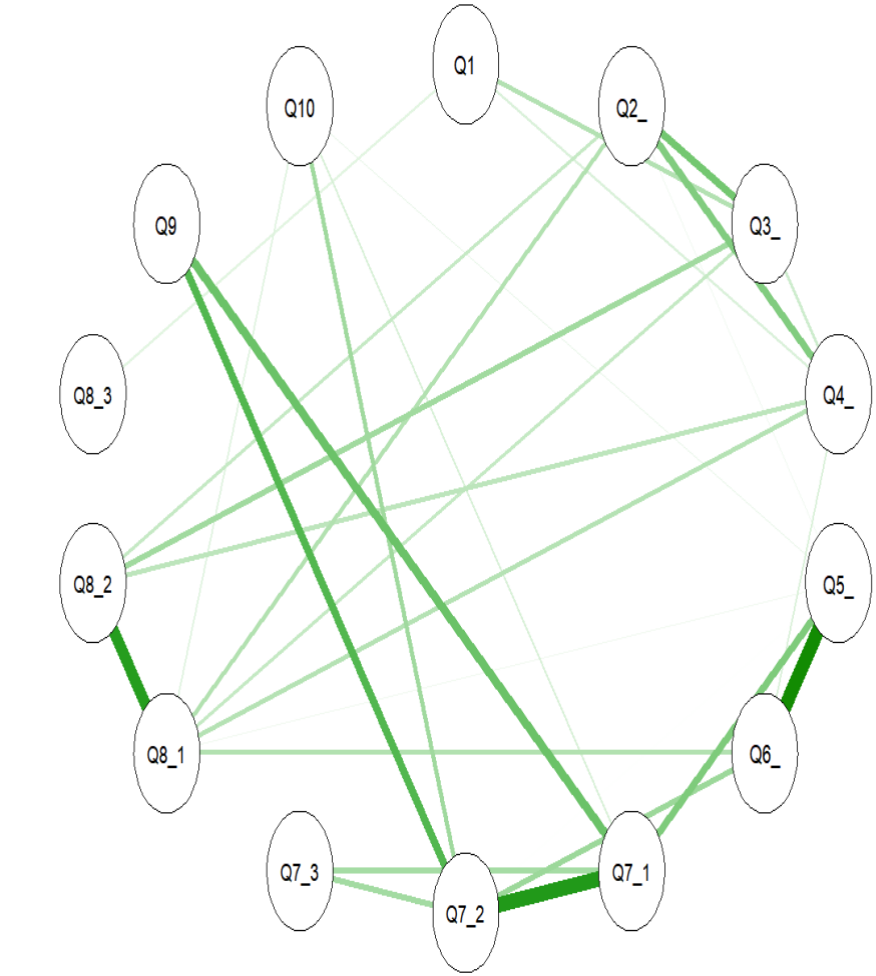
Q1 Trigger	Q7_1 Thought: There is something wrong with me as a person	Q8_3 Reaction: I drink alcohol, take medication or use drugs
Q2 Disturbing & unwanted memories	Q7_2 Thought: I can't rely on myself	Q9 I got no sleep in the last 24 hours
Q3 Nightmares and flashbacks	Q7_3 Thought: I will not be able to control my anger and will do something terrible	Q10 It is hard to hold a thought in my mind for long
Q4 Upset or strong physical reaction	Q7_4 Thought: People can't be trusted	Q11 I feel I can't remember things accurately
Q5 Superalert, watchful or on guard	Q8_1 Reaction: I try hard to control my emotions	
Q6 Jumpy or easily startled	Q8_2 Reaction: I numb my feelings	

Participant 103: Clinician's Formulation

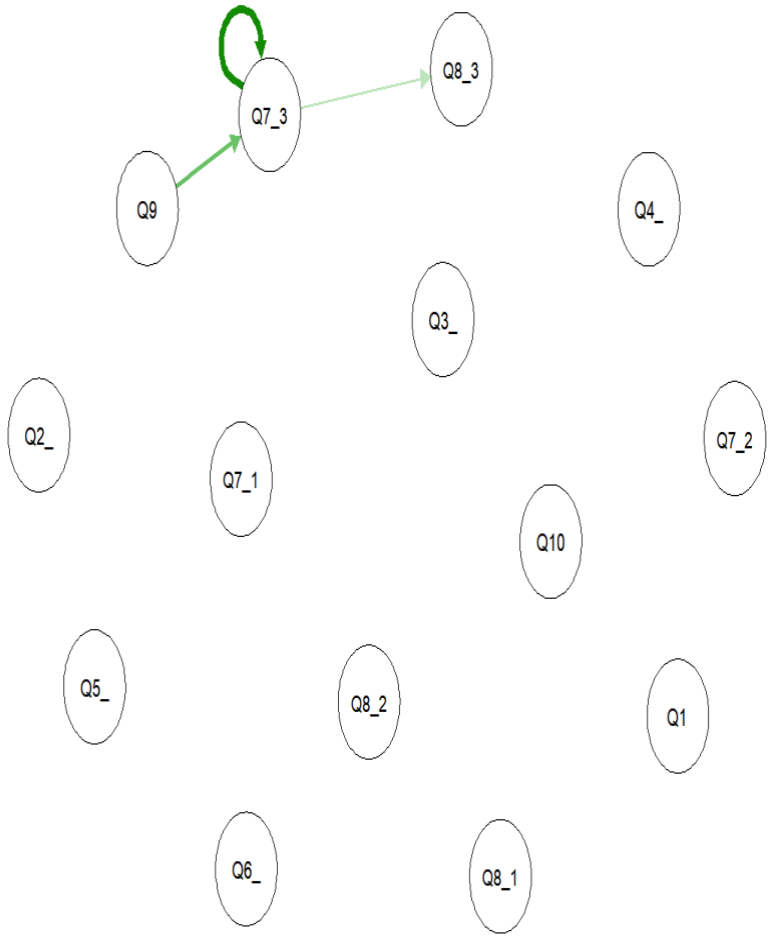
Participant 104: Contemporaneous and Directed Networks

Green = positive correlation Red = negative correlation

Partial Contemporaneous Correlations

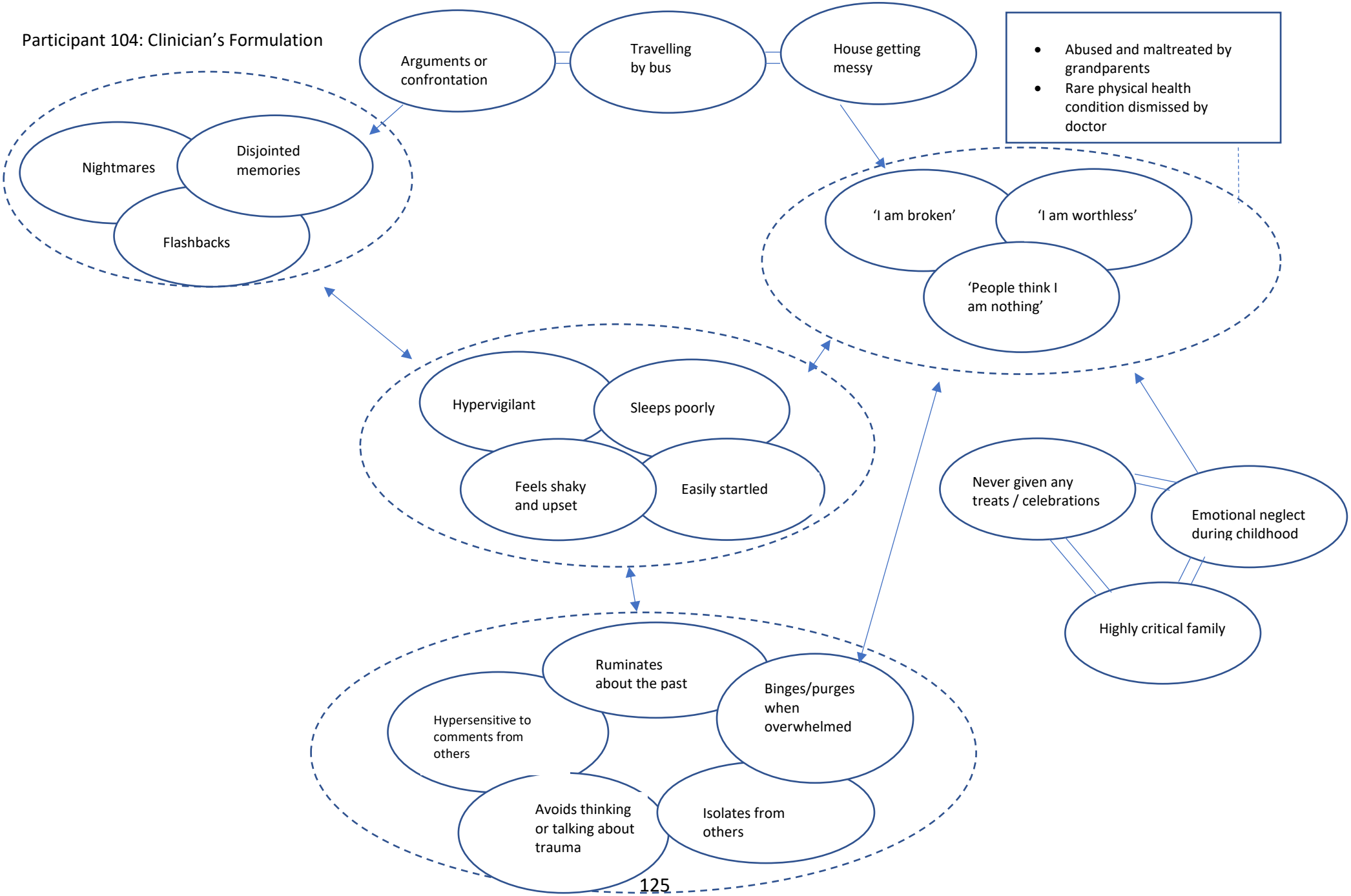


Partial Directed Correlations



Q1 Trigger	Q6 Jumpy or easily startled	Q8_2 Reaction: I go over what happened again and again
Q2 Disturbing & unwanted memories	Q7_1Thought: I have to be on guard all the time	Q8_3 Reaction: I drift off into a world of my own
Q3 Nightmares and flashbacks	Q7_2 Thought: You can never know who will harm you	Q9 Right now I wish I were invisible
Q4 Upset or strong physical reaction	Q7_3 Thought: I will never be able to feel normal emotions again	Q10 Right now I feel mentally robbed
Q5 Superalert, watchful or on guard	Q8_1 Reaction: I dwell on what other people have done to me	

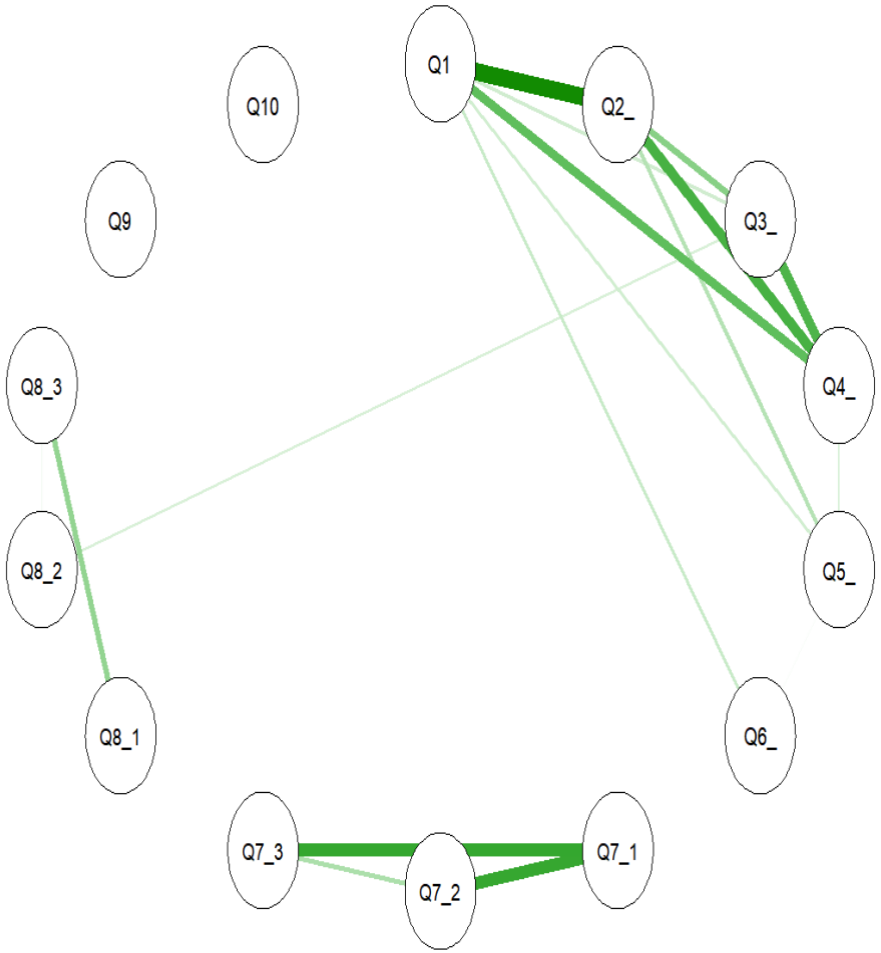
Participant 104: Clinician's Formulation



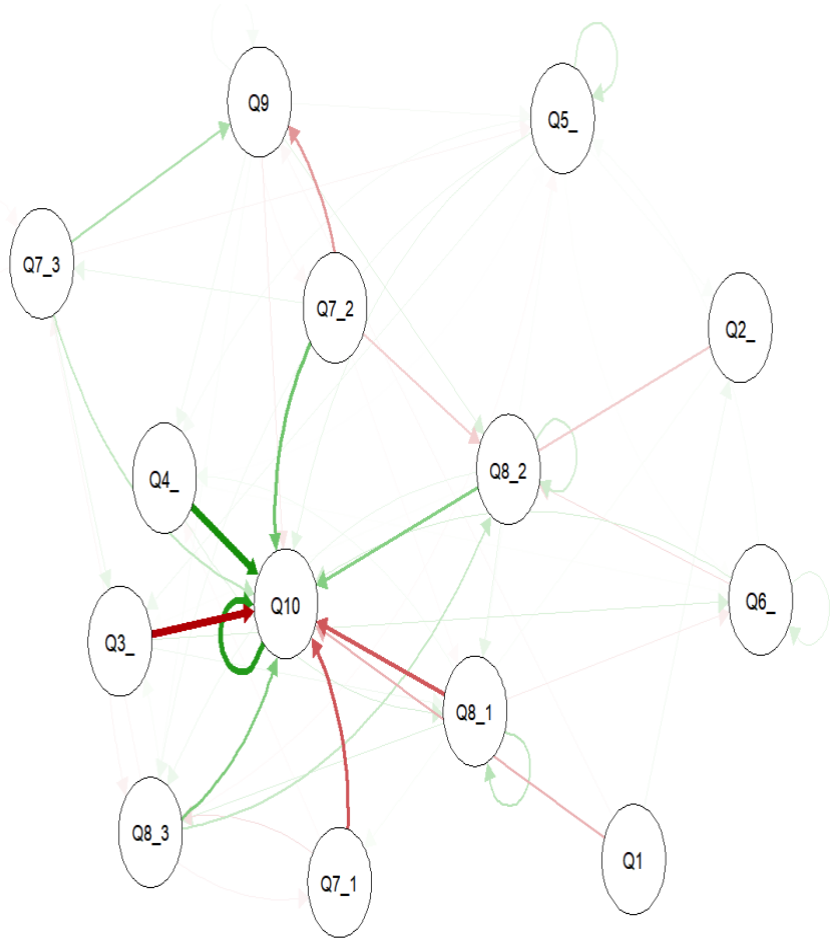
Participant 105: Contemporaneous and Directed Networks

Green = positive correlation Red = negative correlation

Partial Contemporaneous Correlations



Partial Directed Correlations

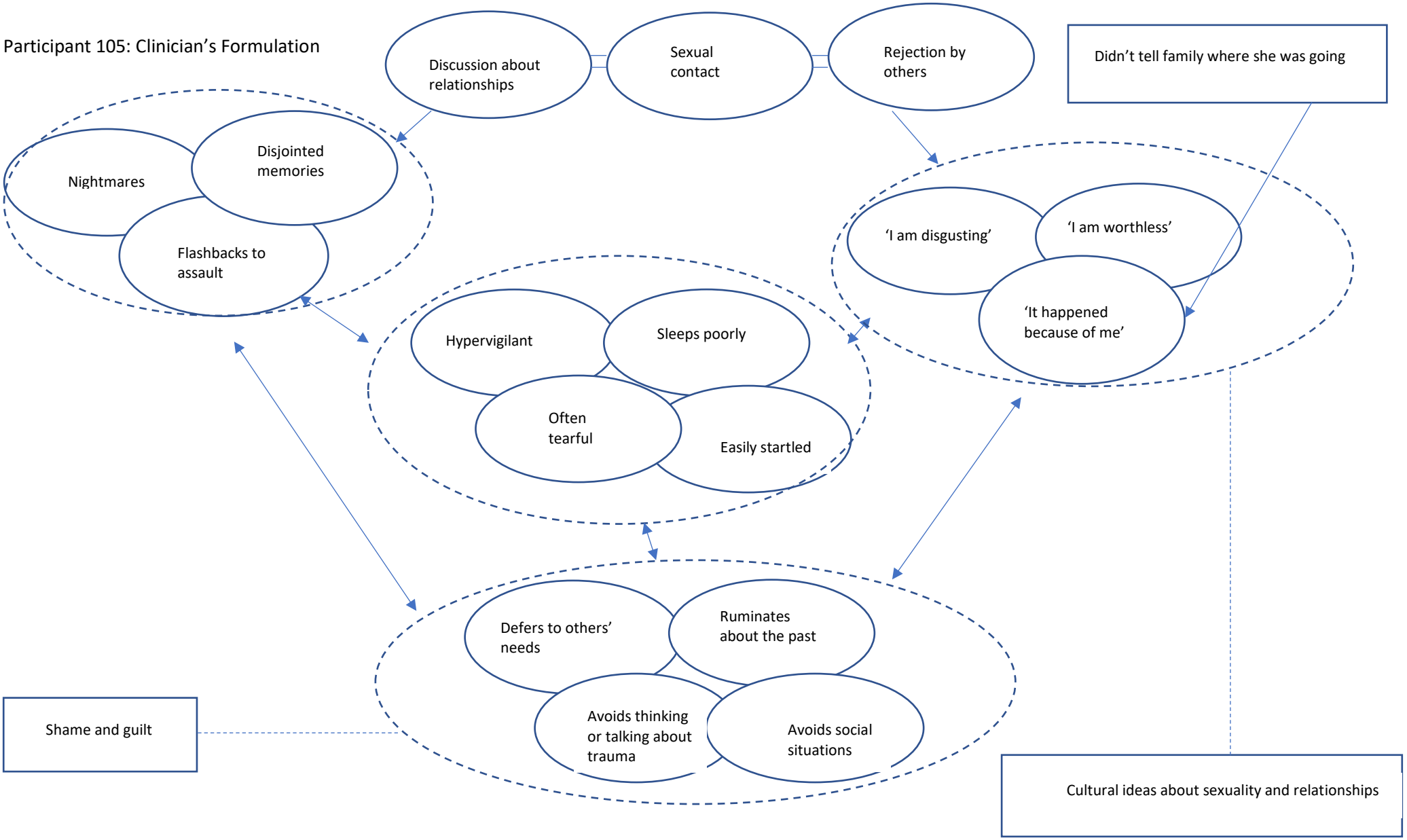


Q1 Trigger
Q2 Disturbing & unwanted memories
Q3 Nightmares and flashbacks
Q4 Upset or strong physical reaction
Q5 Superalert, watchful or on guard
Q6 Jumpy or easily startled

Q7_1 Thought: There is something about me that made the event happen
Q7_2 Thought: I will never be able to feel normal emotions again
Q7_3 Thought: People can't be trusted
Q8_1 Reaction: I work hard at keeping busy with other things
Q8_2 Reaction: I worry that something similar will happen to me or my family
Q8_3 Reaction: I detach myself from the memories

Q9 Right now I am in an academic lecture / seminar
Q10 Right now I have slept less than 3 hours last night
Q11 Right now my sister is out socialising

Participant 105: Clinician's Formulation



Complexity

It is important to acknowledge that some aspects of both approaches to formulation were pre-determined ahead of data collection. For example, for both clinician and network-derived formulations, the type of information included was shaped by the cognitive model of PTSD, although the extent to which this dictated content differed.

In the case of network-derived formulations, content included was heavily pre-determined ahead of data collection. First, not only did the type of information selected as variables reflect the cognitive model, it was tailored to include data about the parts of the cognitive model we were most concerned with investigating, such as cognitions and affective responses. Second, data included was shaped further by researcher choices in respect of the number of variables of each type, for example, the choice to include between three to four cognitive variables. Third, the scope of our network analysis was circumscribed by limits on the total number of variables possible to include, which was determined by consideration of what was required for sufficient insight, but also tolerable to participants and likely to produce accurate data, given the frequency of data collection. As a result, all network-derived formulations followed the same pattern of including between 14-16 elements including six diagnostic features of PTSD (trigger, distressing memory, nightmare/flashback, feeling upset/having a strong physical reaction, feeling 'superalert', feeling jumpy/easily startled), the three or four most representative cognitions from the PTCI (Foa et al., 1999), the three or four most representative affective responses from the RIQ (Clohessy & Ehlers, 1999; Murray, Ehlers, & Mayou, 2002) and two or three contextual moderators.

In the case of the clinicians' formulations, every participant's formulation included content reflecting the five principal elements of the cognitive model (matching triggers, the nature of trauma memories, sense of current threat, cognitive appraisals and dysfunctional behaviours/cognitive strategies) but was more flexible than the network formulations in terms of also permitting information from outside these key categories to be included. Finally, data included in the networks was shaped by being based on self-reports by the participants within the same three-hour window

for the estimated contemporaneous network and within two consecutive three-hour windows of measurement within the estimated directed network. The source of clinicians' data was clinical assessment and interview over the course of therapy, without limit on including material from earlier in the client's life. Having noted these pre-determined structures, Table 11 reflects the findings of our analysis of complexity.

Table 11

Analysis of Formulation Complexity

Participant	Number of formulation elements		Types of information included		Integration	
	Network	Clinician	Network	Clinician	Network	Clinician
101	16	50	Contemporaneous network: • *Diagnostic features of PTSD:* Strong relationships forming a cluster between all variables reflecting diagnostic features of PTSD, except trigger (weakly connected to memories) and flashbacks/ nightmares (weakly connected to memories and feeling upset). • *Cognitions:* Strong inter-variable relationships between first three cognitions relating to feeling unsafe. • *Affective responses*: Strong relationship between two affective responses related to rumination. • *Cognition-affective response links:* Medium strength relationship between cognition 'You can never know who will harm you' and affective response 'I dwell on how I used to be before the event'.	• *Triggers:* Multiple matching triggers including work stressors, being stuck in flat, unwanted male attention. • *Nature of traumatic memories*: Trauma memories were disjointed. • *Appraisals of trauma and its sequelae:* Multiple negative self-appraisals including 'my life is ruined', 'my family blames me' and 'I am weak/damaged' which are also shaped by ambivalent attachment to perpetrator of trauma. • *Sense of current threat* is exacerbated by feeling guilty and ashamed. • *Biological factors:* Nightmares are influenced by history of sleep paralysis. • *Core beliefs* such as feeling they should keep their needs to themselves, which were	Clear representation of correlations between observed symptoms within/over a specific time frame without integration of these elements into a 'meaningful account'. Noted co-occurrence of several diagnostic symptoms of PTSD (memories, feeling upset, feeling superalert and jumpy) without trigger, suggesting symptoms may be more driven by rumination than triggers. Co-	Diagram links elements to constitute a comprehensive, meaningful explanation of symptom interactions in terms of why and how they relate, how difficulties evolved and are maintained, especially in respect of pre-disposing factors, their role in forming core beliefs about the need for self-reliance and being successful and the way these beliefs and feelings of shame currently

Participant	Number of formulation elements		Types of information included		Integration	
	Network	Clinician	Network	Clinician	Network	Clinician
			• *Affective response-contextual moderator links:* Negative correlation between feeling isolated and contextual moderator of being with friends. Directed network: • *Contextual moderator-affective response links:* One positive correlation between contextual moderator 'I have had an alcoholic drink today' and affective response 'I drink alcohol, take medication or used drugs'.	driven by pre-disposing factors such as childhood experiences. • *Effect of trauma on core beliefs which then influence cognitions:* The shattering by the trauma of their sense of personal capability interacted with core beliefs to drive negative self-appraisals. • *Contextual factors* such as social & financial stress worsens dysfunctional coping, which then leads to rumination and further dysfunctional coping in a negative spiral.	occurrence of multiple cognitions plus co-occurrence of affective responses relating to rumination reinforces impression of rumination driving symptoms. Clinician interpretation required to link observations meaningfully.	interact with dysfunctional coping strategies such as rumination to maintain symptoms.
102	16	37	Contemporaneous network: • *Diagnostic features of PTSD:* Strong relationships between triggers, nightmares/ flashbacks & feeling upset. • Strong relationship between feeling 'superalert' and jumpy. • *Diagnostic features of PTSD/contextual moderator link:* Strong negative correlation between PTSD symptom 'feeling upset' and	• *Triggers:* Clear, specific triggers which lead to flashbacks/nightmares and negative self-appraisals. • *Appraisals of trauma and its sequelae:* Negative self-appraisals related to trauma link to self-blame. • *Dysfunctional behaviours/cognitive strategies:* Includes avoidance of social contact, thoughts, drug use, restricted eating.	Network confirmed co-occurrence of diagnostic symptoms of PTSD as well as co-occurrence of dysfunctional eating with these symptoms. Limited broader understanding of participant's	Detailed elaboration of multiple aspects of presentation, presented in a way which allows current difficulties to be understood – inferred information such as explanations relating to feelings of shame, guilt,

Participant	Number of formulation elements		Types of information included		Integration	
	Network	Clinician	Network	Clinician	Network	Clinician
			contextual moderator of eating a normal amount of food. • *Affective responses:* Positive correlation between having a flashback/nightmare and numbing feelings. • *Cognitions:* Weak inter-variable relationships between cognitions about feeling unsafe. • *Diagnostic features of PTSD/cognition link:* Positive correlation between feeling upset and thinking 'you can never know who will harm you'. No directed network.	• *Pre-disposing factors* for self-blame include being bullied and critical parenting. • *Historical factors* influencing dysfunctional ways of coping include history of self-harm and disordered eating. • *Explanations:* Overly developed sense of responsibility is an explanation for self-blame and sense of guilt, shame and defectiveness. • Shame and guilt drive dysfunctional coping which then re-triggers these emotions in negative spiral.	experiences obtained.	self-blame and their interactions with appraisals of the traumatic event and subsequent cognitive strategies and dysfunctional behaviours assists this insight.
103	16	33	Contemporaneous network: • *Diagnostic features of PTSD:* Strong positive correlations between triggers, memories and flashbacks. • Strong positive correlations between feeling upset and flashbacks and triggers. • *Diagnostic features of PTSD/affective response link:* Strong positive correlations	• *Appraisals of trauma and its sequelae:* Multiple negative self-appraisals in relation to the traumatic event also shaped by previous repeated interpersonal trauma. • *Nature of trauma memory:* Fragmented nature in part due to peri-traumatic dissociation. • *Pre-disposing factors* including childhood neglect and	Network confirmed strong co-occurrence of several diagnostic features of PTSD together with affective response to drink alcohol, take medication or drugs, suggesting this is a coping	Well integrated to provide explanation of how expression of current difficulties may be influenced by historical difficulties with emotional avoidance arising out of childhood

Participant	Number of formulation elements		Types of information included		Integration	
	Network	Clinician	Network	Clinician	Network	Clinician
			between flashbacks and affective response to drink alcohol, take medication or drugs. • *Affective responses:* Medium strength correlation between numbing feelings and drinking alcohol, taking medication or drugs. • *Diagnostic features of PTSD/affective response link:* Weak positive correlation between feeling 'superalert', and affective response to drink alcohol, take medication or drugs. • *Diagnostic feature of PTSD/cognition link:* Weak positive correlation between memories and the thought that people can't be trusted. Directed network: • Autocorrelation 'I got no sleep in the last 24 hours'.	emotional abuse and difficulties with friendships seem to drive difficulties with self-soothing which then affect nature of unhelpful cognitive and behavioural strategies used. • *Explanations:* Several of the unhelpful behaviours / cognitive strategies relate to avoidance of emotions which is explained by difficulties in regulating emotions and self-soothing. • Self-harm explained as method of making self feel better. • Shame, guilt, disgust and anger arising from the trauma maintain the sense of current threat.	mechanism. Co-occurrence of use of drugs/alcohol/ meds with affective response 'I try to numb my feelings' supports this interpretation. Co-occurrence helpful in indicating links but limited further explanation of experiences.	neglect and how current dysfunctional coping strategies may also be shaped by emotional avoidance and difficulties with self-soothing.
104	14	19	Contemporaneous network: • *Cognitions/ contextual moderator link:* Strong positive correlations between the	• *Appraisals of the trauma and its sequelae*: All negative self-appraisals such as 'I'm	Several clusters of strong correlations suggested a strong sense of current	Clinician supplied information for inclusion within diagram

Participant	Number of formulation elements		Types of information included		Integration	
	Network	Clinician	Network	Clinician	Network	Clinician
			thoughts 'I have to be on guard all the time' and 'You can never know who will harm you' and contextual moderator 'Right now, I wish I were invisible'. • *Diagnostic features of PTSD:* Strong positive correlations between feeling 'superalert' and being jumpy. • *Affective responses:* Strong positive correlations suggesting rumination between affective responses 'I dwell on what other people have done to me' and 'I go over what happened again and again'. • *Diagnostic features of PTSD/affective response link:* Multiple medium-strength positive correlations between symptoms of PTSD (memories, flashbacks/ nightmares and feeling upset) and affective responses 'I dwell on what other people have done to me' and 'I go over what happened again and again'. Directed network:	worthless' and 'I have a label that says 'abuse me''. • *Triggers:* Diverse matching triggers including confrontation and moving house. • *Biological information*: Participant had a rare physical health condition which was diagnosed late by GP. Experience of diagnosis confirmed negative self-appraisals as GP was initially dismissive. • *Historical pre-disposing factors* including being subjected to physical violence and degradation prior to the trauma influenced the negative self-appraisals of the trauma and its sequelae. • *Sense of current threat* expressed as tearfulness and anxiety. • *Explanation* in terms of previous interpersonal experiences offered for dysfunctional behaviours and coping strategies.	threat, feeling jumpy and ruminating on the traumatic event were notable elements of the participant's presentation. Co-occurrence of variables were descriptive of participants' experience and provided some level of explanation but this was limited and hard to pull together into a more comprehensive understanding of what drove and maintained daily symptomatology.	framework. With framework information is a meaningful explanation of difficulties, though less elaborated than for other participants. Inclusion of historical pre-disposing factors such as previous abuse provide insight into appraisals of trauma relating to feelings of worthlessness and suggest core beliefs which may contribute to maintaining difficulties.

Participant	Number of formulation elements		Types of information included		Integration	
	Network	Clinician	Network	Clinician	Network	Clinician
			• *Cognitions/contextual moderator link:* Thought 'I will never be able to feel normal emotions again' predicted by 'Right now, I feel invisible'. • *Affective response* 'I drift off into a world of my own' predicted by 'I will never be able to feel normal emotions again'. • Autocorrelation 'I will never be able to feel normal emotions again'.	• Traumatic memories appear as intrusions and are also maintained through rumination.		
105	15	15	Contemporaneous network: • *Diagnostic features of PTSD:* Strong positive correlations forming a cluster between trigger, memories, flashbacks/ nightmares and feeling upset. • *Cognitions:* Strong positive correlations between 'There is something about me that made the event happen', 'I will never be able to feel normal emotions again' and 'People can't be trusted'. • *Affective responses:* Medium strength positive correlation between avoidant affective	• *Triggers:* Emotional states including feeling rejected by others romantically and interpersonal events such as particular types of conversation. • *Context for specific traumatic event* is provided which helps explain how the negative self-appraisals resulting from the traumatic event developed. • *Historical and cultural factors* offered which may also have shaped interpretation of traumatic event e.g. cultural ideas about virginity.	Cluster of co-occurring diagnostic features of PTSD (trigger, memory, flashback, feeling upset) confirms presence of these symptoms. Strong co-occurrence of cognitions about lack of trust in others, self-blame and not feeling 'normal' emotions, without	Formulation was brief, coherent and used the descriptive information to infer explanations including that the need to avoid upsetting others maintains the idea of her own worthlessness. Together with inclusion of elements such as cultural ideas

Participant	Number of formulation elements		Types of information included		Integration	
	Network	Clinician	Network	Clinician	Network	Clinician
			responses 'I work hard at keeping busy with other things' and 'I detach myself from the memories'. Directed network: • *Contextual moderators:* Multiple positive and negative correlations with many variables both positively and negatively connected to the contextual moderator 'Right now, I have slept less than 3 hours last night'.	• *Explanation:* Feelings of shame and worthlessness offered as underlying explanation for dysfunctional behaviours and coping strategies. • *Maintenance:* Avoidance of upsetting others noted as a coping strategy which maintains other symptoms as it means participant's sense of worthlessness is unchallenged.	connection to PTSD symptoms suggests these may be separately occurring rumination. A further separate link between affective responses relating to avoidance suggest this is also a strategy used, but overall little comprehensive explanation of what maintains symptoms.	about virginity this helped create an integrated explanation of difficulties and their meaning for the participant, as well as precipitating and maintaining factors.

Given the limits already noted, our finding that a greater number of elements were included in the clinician-derived formulations overall was not surprising; (clinician: *m* 30.8, SD = 12.66, range = 15-50; networks: *m* = 12.8, SD = 2.70, range = 14-16). However, further analysis revealed that relying solely on the mean score for the group of five clinician-derived formulations obscured a possibly bimodal distribution in the results, with a notably higher mean number of formulation elements included in formulations 101-103 than in formulations 104-105 (101-103: *m* 40, SD =7.26, range 33-50; 104-105: *m* 17, SD = 2, range = 15-19). Although this did not change the finding that a greater number of elements were included in the clinician-derived formulations, it suggested that factors other than the formulation approach may influence this particular aspect of formulation.

Comparison of the networks and clinicians' formulations confirmed that both included pre-determined elements which reflected the cognitive model of PTSD, such as triggers, cognitions and affective responses. As predicted, the clinicians' formulations also included a broader range of types of information, for example, core beliefs about needing to be self-sufficient and capable for participant 101, historical pre-disposing factors such as childhood neglect for participant 103 and explanations such as the overly developed sense of responsibility which maintained symptoms for participant 102. These exemplify the range of additional idiographic information, especially inferred material, which was included within clinicians' formulations in contrast to the network formulations, which were limited to descriptive content. The variety and volume of idiographic information added to clinicians' formulations suggested that it may be unrealistic to expect network formulations to capture the same detail and complexity using a 'one size fits all' fixed structure of variables.

In order to unpick what may have contributed to differences between the formulations for each participant, we considered whether a sufficient sample of each type of variable had been included. It is difficult to predict with confidence the optimum number of each type of variable required to capture a particular participant's idiosyncratic difficulties and yet the process of network analysis requires variables to be defined in advance. Even so, our comparison suggested that limitations in the network formulations did not arise principally from an insufficient sample of each

type of variable, but rather from the variables used being insufficiently precise and personalised. For participant 101, the strong co-occurrence of multiple cognitive variables including the self-blaming cognition 'The event happened to me because of the sort of person I am' as well as the presence of co-occurring affective variables 'I dwell on how I used to be before the event' and 'I dwell on what other people have done to me' within the contemporaneous network conveyed a powerful sense of rumination and self-blame, which was also identified by the clinician's formulation. However, the network formulation did not identify the role of specific thoughts about needing to be independent and successful and the consequences of failing in this as an additional driver of this participant's symptoms, as identified in the clinician's formulation. These idiosyncratic cognitions were, unsurprisingly, not included in the PTCI (Foa et al., 1999) from which the cognitions for inclusion in the network analysis were selected. Including additional cognitions which were still insufficiently tailored to this participant would not have improved the network's capacity to capture this aspect of the participant's difficulties.

Reinforcing this point, the severity of participant 102's sense of current threat was evident in both the network and clinician's formulation from the strong correlation between being superalert/on guard and jumpy in the network formulation and the number of behaviours suggesting a sense of threat in the clinician's formulation. However, only the clinician's formulation included the peri-traumatic cognition 'he's going to kill me' as an appraisal directly contributing to this sense of threat. As this type of thought would have occurred during rather than following the traumatic event, it was not included within the PTCI (Foa et al., 1999) and so not available for the participant to select for inclusion as a network variable, despite its likely relevance. In both these cases, while including a larger sample of cognitive variables may have assisted in capturing additional cognitions which contributed to difficulties, our comparison suggested that using variables which were more individually tailored to that participant, rather than taken from a standardised measure, would have allowed the network formulation to reflect the participant's experience more accurately.

In terms of the affective responses included, we found a similar pattern where limitations in what was captured by the network appeared to arise from the network variables chosen being less specific in identifying what maintained symptoms than the clinician's formulations. Participants' network variables often included rumination (participant 101), distraction (participant 105) and avoidance (participant 103) without identifying the emotions those participants were trying to avoid. In contrast, the clinicians' formulations specified feelings of shame (participant 101, 103, 105), guilt (participant 101, 103) and self-blame (participant 105) as important underlying drivers of symptomatology, with symptoms then maintained by dysfunctional strategies such as rumination, distraction and avoidance. The inclusion of additional variables capturing feelings such as shame and guilt may have allowed the network to represent participant's emotional reactions in a way that was more comparable to the clinicians' formulations, but these variables were not available within the RIQ (Clohessy & Ehlers, 1999; Murray, Ehlers & Mayou, 2002), which was the source material used for the affective responses in our network. Instead variables within the RIQ (Clohessy & Ehlers, 1999; Murray, Ehlers & Mayou, 2002) were principally indirect measures of emotions, which focused largely on behavioural responses.

Participants were given more flexibility in their selection of contextual moderators, and allowed to include anything they felt affected their symptoms. As a result, the range of contextual moderators included was broad and the specific moderators more precisely tailored to each participant (sleep patterns for participant 105, alcohol/drug consumption for participant 101) and thoughts ('I feel mentally robbed' for participant 104, 'my sister is out socialising' for participant 105). The sample of contextual moderators included did not appear to be insufficient, although it may have been useful to define moderators more accurately; there was often overlap between networks' contextual moderators and the clinicians' formulation (alcohol/drug use for participant 101; dysfunctional eating for participant 102), although within the clinician's formulations the same factors were often more properly defined as dysfunctional behaviours/emotional responses rather than symptom moderators.

The participants whose networks appeared to provide the least complete explanation of the symptoms in comparison to the clinicians' formulations were those whose clinician's formulation relied most heavily on inferred material; an unsurprising finding given that this material was not included in the networks. The network formulation for participant 103 was particularly sparse, as the important historical pre-disposing factor of childhood neglect leading to current difficulties with emotional regulation and self-soothing included in the clinician's formulation was entirely absent from the network. In terms of the approach used in this study where network variables were largely selected by participants, the degree to which difficulties such as managing emotional regulation are possible to capture within a network is likely to be determined by the participant's degree of self-insight in relation to that difficulty as well as their willingness to monitor it. Calibrating the degree to which you are not achieving something is a difficult and potentially distressing activity in comparison to recording the presence or absence of a particular symptom. As a result, this type of emotionally aversive and inferred variable may be hard to include in networks reliant on participants identifying relevant variables.

In terms of formulation integration, the inclusion of this kind of inferred material and the way it was used to link and explain clinical observations was observed to be a key contributor to the clinicians' formulations having greater explanatory power and a better sense of integration than the networks. As noted, the absence of inferred material in the network for participant 103 contributed to the limitations of that network in providing insight into the participant's difficulties. Similarly, for participant 102, explanations provided by the clinician in relation to the participant's overly developed sense of responsibility as well as their sense of shame, guilt and defectiveness gave helpful insight into the factors which shaped their appraisals of the traumatic event and maintained their dysfunctional coping. Without this inferred information, participant 102's contemporaneous network captured the co-occurrence of diagnostic symptoms of PTSD with maladaptive eating patterns, but did not go further to explain how this participant's particular presentation might have

developed, what sustained it and so, by extension, what might help to reduce their difficulties, all of which was provided by the clinician's formulation.

In summary, our findings were that there were substantial differences between clinician- and network-derived formulations in terms of the number of formulation elements included, the types of information included and the degree to which this information was integrated into a cohesive picture of the person's difficulties. Therefore, according to the definition of complexity used in this analysis, the part of the hypothesis stating that network analysis-derived formulations will resemble therapist-derived formulations in terms of complexity was not supported.

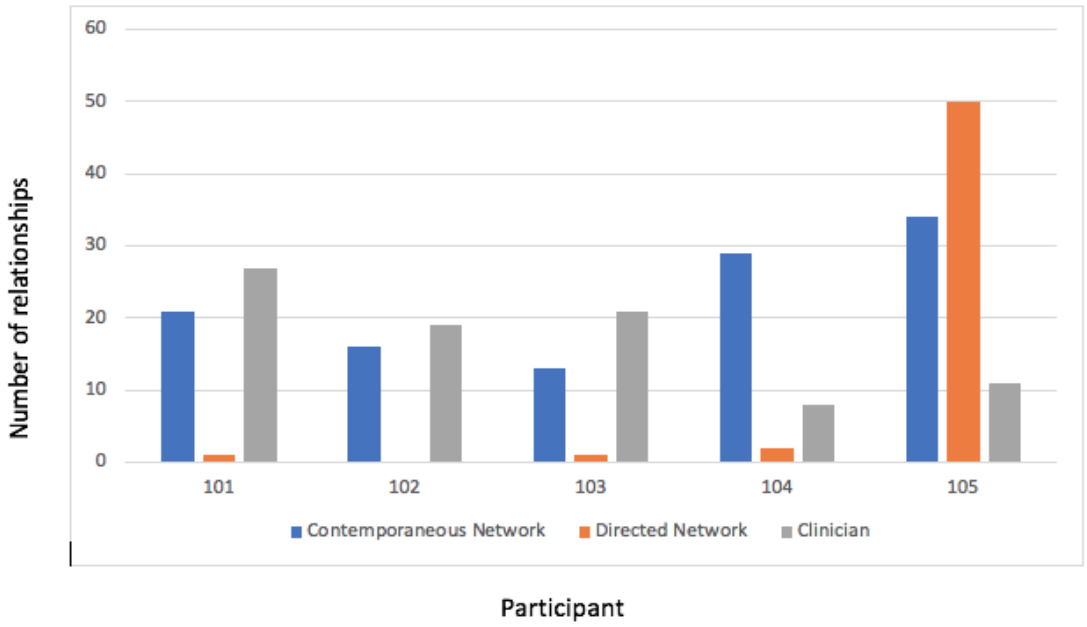
Relationships between Elements of the Problem

Analysis of the relationships between elements of the problem focused on the number of relationships within each formulation and the degree of similarity of the relationships between variables when two formulations for the same participant were compared.

Number of Relationships within the Formulation. The number of relationships included in each type of formulation are reported in Figure 3.

Figure 3

Number of Inter-Symptom Relationships Identified by Type of Formulation



No clear pattern in terms of number of relationships was observed. For participants 101-103 more relationships were included in the clinician-derived formulation by a small margin. For participants 104 and 105 notably more relationships between variables were included in network-derived formulations. The degree of difference in the number of relationships identified by clinicians and networks for participants 104 and 105, particularly in the network of directed symptoms for participant 105, suggested other factors may be influencing the results of these particular comparisons. One possibility was the use of the relatively loose LASSO regularization in estimation of the networks, which may have produced a large number of false positive relationships in the case of Participant 105's directed network. This finding indicated that more accurate estimation of networks may be produced by setting gamma on an individual basis for each participant. Criteria by which to determine the appropriate level of gamma for a participant are as yet unclear, although verification of the credibility of the network produced using different levels of gamma by comparing it with an alternative, such as a clinician's formulation, may be one method to achieve this.

Similarity of Connections. In order to establish how far relationships within the clinicians' and network-derived formulations resemble one another, it is important to clarify first the kind of relationships between variables included within the two approaches.

Within the clinicians' formulations, solid lines connecting variables indicated causal relationships between variables, with an arrowhead indicating the direction of causality. Explanations were connected to the areas of the formulation to which they applied via dotted lines. All variables within pre-determined variable categories, for example, 'sense of current threat', were inter-linked either by double parallel lines, or by overlapping variable labels; both indicated that these grouped variables were considered to have equivalent relationships to all other variables within the formulation. All relationships within a clinician's formulation are considered to be hypothetical, though based on clinical observation and assessment.

Within network formulations, lines connecting two variables, known as 'edges', indicated partial correlation between those variables; in other words, that those variables had been found to

co-occur within the same window of measurement for a contemporaneous network, or in the case of a directed network, that one variable had been found to predict another in consecutive windows of measurement, with an arrowhead indicating the direction of effect. Relationships between variables in network formulations were based on empirical data. The thickness of the 'edge' indicated the strength of correlation found between variables.

Visual analysis of the two formulation approaches indicated significant areas of overlap in terms of the relationships between symptoms they depicted, including identifying idiographic relationships such as the connection between experiencing symptoms of PTSD and eating in a dysfunctional way for participant 102, and between flashbacks and rumination for participant 104. However, this comparison also indicated a number of differences in terms of the relationships included. As noted previously, the types of variables included within both network- and clinician-derived formulations were predetermined to some extent by the cognitive model of PTSD. Within the clinician-derived formulation, relationships between the key components of the formulation (matching triggers, trauma memories, sense of current threat, appraisals and dysfunctional behaviours/cognitive strategies) were also pre-determined by the cognitive model, with fixed reciprocal causal relationships assumed to exist between all five key components. When clinicians added idiographic elements to this model, these additional causal relationships were idiographic, reflecting the clinicians' understanding of that participant's unique presentation. In contrast, within the network model, while variables included were also pre-determined by the cognitive model, edges included were always idiographic, being based on what was found in the empirical data. For some participants, the inclusion of pre-determined, assumed relationships between some elements in the clinician's formulation may explain differences found between relationships in the network and clinician-derived formulations. The contemporaneous networks estimated for participants 103, 104 and 105 did not show a strong correlation between hypervigilance (Q5) and other diagnostic symptoms of PTSD (Q1-4) although this is part of the cognitive model of PTSD and therefore causal relationships between hypervigilance/sense of current threat and the other key symptoms of PTSD

were included in the clinicians' formulations for these participants. While it cannot be entirely ruled out that the networks failed to identify these relationships, this difference may indicate the network formulations being more idiographic and so accurate in their representation of inter-variable relationships in this instance, which meant it did not include relationships between hypervigilance and other symptoms not supported by the data, despite those being included in the cognitive model. It should be noted that the network formulations often did find relationships between key components of the diagnostic criteria for PTSD, which overlapped to a large degree with the cognitive model, but that the exact 'constellation' of symptoms found to co-occur for each participant showed individual differences which also supported the idea of the network formulations being more accurate and idiographic than clinicians' formulations in respect of relationships found between core components of the cognitive model.

A further difference between relationships in the clinicians' and network-derived formulations was found in examples of variables identified as important and strongly inter-connected to other variables via multiple causal links in the clinician's formulation, but lacking any relationships to other variables in either network formulation. The clinician's formulation for participant 103 identified the participant's shame, guilt and difficulties with self-soothing as important causal drivers of their symptomatology. Therefore, it may have seemed reasonable to expect that variables reflecting these feelings and selected by participant 103 for inclusion as relevant cognitions would have been found to have multiple connections to other variables, given that central nodes within network analyses are highly interconnected. However, the networks for participant 103 including variable 7_1 'There is something wrong with me as a person' and variable 7_3 'I will not be able to control my anger and will do something terrible' were estimated to contain no edges at all connecting these variables to others within either the contemporaneous or directed network for participant 103. Examination of the raw data confirmed that both variables were consistently rated very highly by the participant for the strength of their belief in them; variable 7_1 'There is something wrong with me as a person' in particular, was almost exclusively given the

highest possible rating for strength of belief. These consistently high ratings may indicate ceiling effects, where these symptoms' variance was restricted by an insufficiently broad range of ratings, meaning that any co-variance with other variables could not be captured and so relationships between these variables and others were not apparent in the network. An alternative explanation is that these variables were more stable than other symptoms, and so varied on a different time scale than could be captured by the window of measurement used, which again resulted in relationships with other variables not being captured. Both possibilities are potential flaws in the network approach. Although the ratings given confirmed that both variables were considered meaningful and true of them by the participant, the lack of relationships depicted within the network formulation suggested they were not significant drivers of symptoms, whereas the clinician's formulation suggested the opposite. This example highlights the specificity of what is identified by relationships within the network formulation, which is correlation and co-occurrence within a particular time frame, and that this specificity may risk missing the importance of variables which are strongly felt by the participant, do not necessarily co-occur with other variables within the interval of measurement but are instead consistently present at a high level. These types of variables may affect other variables without that effect being registered at all within the parameters set for relationships within network formulations, although their significance was identified within the more flexible context of clinicians' formulations, creating an important difference in the relationships included.

Given the differences found in the number and especially the type of relationships included in the two formulation approaches, our findings indicated that, despite some overlap, network analysis-derived formulations and therapist-derived formulations only partially resembled one another in terms of relationships between elements of the problem and that therefore, hypothesis one was not supported.

Hypothesis 2

This study aimed to operationalise within a network study the theoretical relationship between cognitive and affect variables at the heart of the cognitive model of PTSD (Ehlers & Clark, 2000). Hypothesis two stated that there would be a 'lead-lag' relationship between cognitive and affect variables. By identifying temporal relationships between cognitive and affect variables within participants' temporal networks, we hoped to understand the extent to which this central relationship had been detected. Direction of effect between cognitive and affect variables was not stated in the hypothesis, as this was an exploratory study and by maximising the number of relationships included we hoped not to miss relevant data.

The cognitive variables for each participant were the three (participants 104 and 105) or four (participants 101, 102 and 103) cognitions selected by the participant as most characteristic of them from their responses to the PTCI (Foa et al., 1999). The affect variables were the three (participant 101, 103, 104 and 105) or four (participant 102) affective reactions selected by the participant as most characteristic of them from their responses to the RIQ (Clohessy & Ehlers, 1999; Murray, Ehlers & Mayou, 2002).

Analysis of Directed Networks

The networks of partial directed correlations are depicted for each participant on the right side in Figure 2. Within the networks, cognitive variables start with 'Q7' and affect variables with 'Q8'.

Findings from the analysis of temporal networks were that no relationships between cognitive and affective variables were included in the estimated network for participants 101, 102, 103 or 104. Four cognitive-affective relationships were included in the temporal network for participant 105 and are depicted in Table 12. One was a positive correlation, meaning the second variable was predicted to occur in the next window of measurement by the first variable; three were negative correlations, meaning that the second variable was predicated not to occur in the next window of measurement by the first variable.

Table 12

Cognitive-Affective Variable Relationships within Temporal Networks

Participant	First window of measurement	Second window of measurement
105	Q7_2 Thought: I will never be able to feel normal emotions again	Q8_3 Affective response: I detach myself from the memories (positive)
	Q7_1 Thought: There is something about me that made the event happen	Q8_3 Affective Response: I detach myself from the memories (negative)
	Q7_2 Thought: I will never be able to feel normal emotions again	Q8_1 Affective Response: I worry that something similar will happen to me or my family (negative)
	Q8_3 Affective response: I detach myself from the memories	Q7_3 Thought: People can't be trusted (negative)

Analysis of Contemporaneous Networks

Directed networks identify only those relationships in which changes in one variable are predicted by changes in another variable in the preceding window of measurement. Relationships in which changes in one variable do predict another but those changes both occur within the same window of measurement appear as partial correlations within the contemporaneous rather than directed network as their direction of effect cannot be determined. In order to ensure all possible relationships relevant to our hypothesis were considered, we identified the cognitive-affect variable relationships within participants' contemporaneous networks, as well as those within the directed network.

The networks of partial contemporaneous correlations are depicted for each participant on the left side of Figure 2. As for the directed networks, cognitive variable labels start with 'Q7' and affect variables with 'Q8'.

Findings were that no cognitive-affect variable relationships were included in the estimated networks for participants 102, 103, 104 and 105. Three positive partial correlations between

cognitive and affect variables were included in the contemporaneous network for participant 101 and are listed in Table 13.

Table 13

Cognitive-Affective Variable Correlations within Contemporaneous Networks

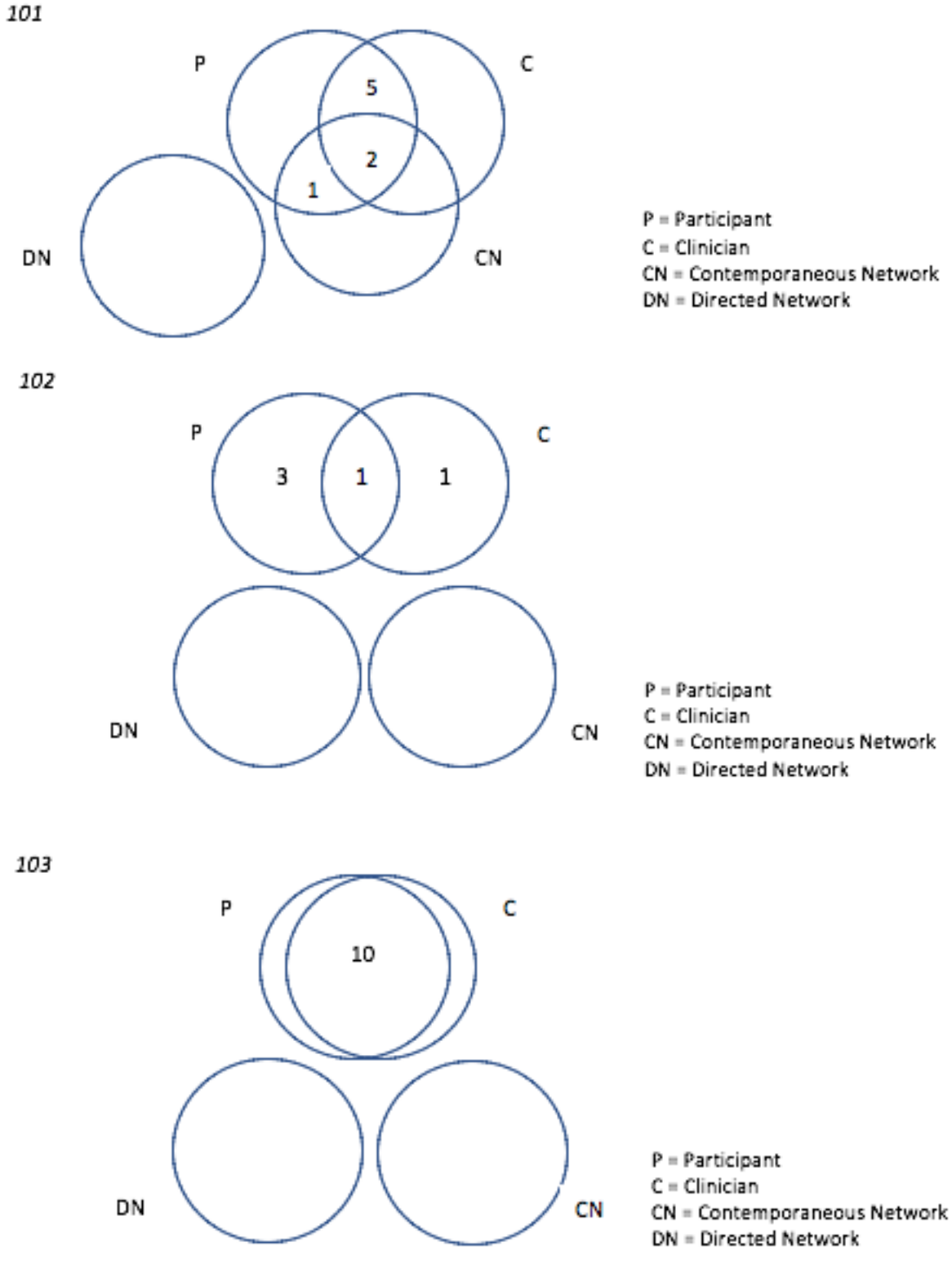
Participant		
101	Q7_2 Thought: You can never know who will harm you	Q8_1 Affective Response: I dwell on how I used to be before the event
	Q7_4 Thought: I feel isolated and set apart from others	Q8_1 Affective Response: I dwell on how I used to be before the event
	Q7_4 Thought: I feel isolated and set apart from others	Q8_2 Affective Response: I dwell on what other people have done to me

Analysis of Perceived Relationships

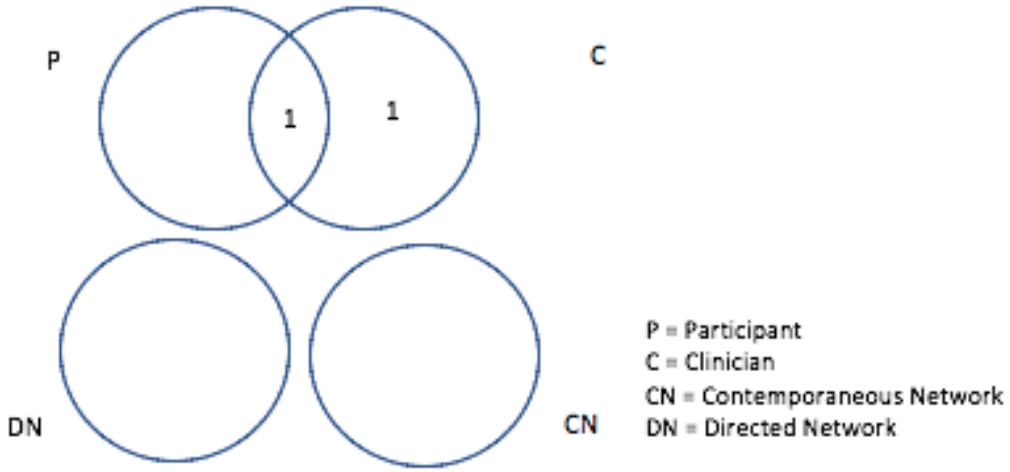
Ahead of data collection, participants and clinicians had been asked to note any inter-symptom relationships they perceived to exist between symptoms included in the participant's personalised survey. The degree of correspondence between relationships between cognitive and affect variables perceived by participants and clinicians and those estimated within the contemporaneous and directed networks is reported in Figure 4. Numbers indicate the number of relationships for which there was agreement.

Figure 4

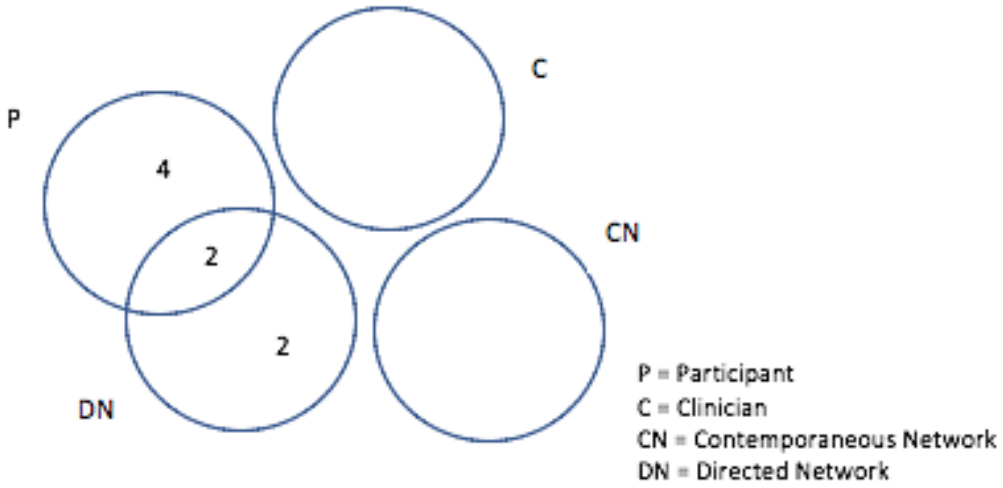
Correspondence between Participant and Clinician Perceptions of Cognitive and Affective Symptom Relationships and Networks



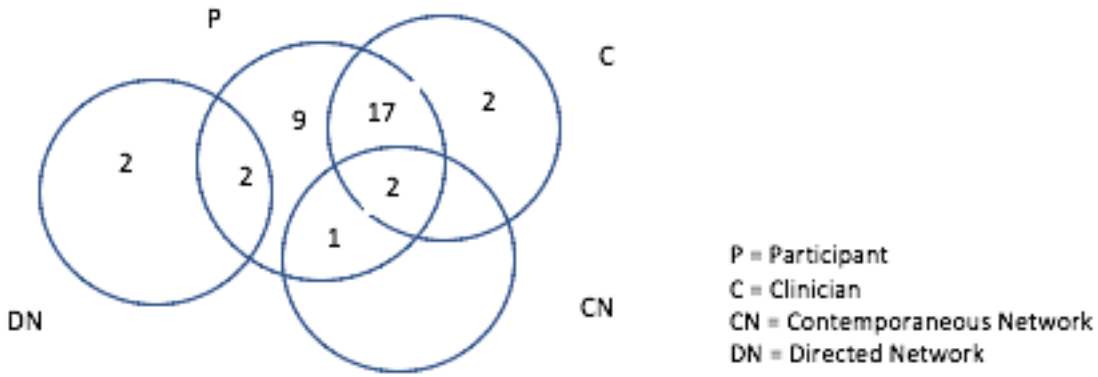
104



105



Total



Our findings were that there was a notable difference between the relationships between specific cognitive and affect variables perceived to exist by the participant and clinician and those included in either the contemporaneous or directed networks. Out of a total of 35 relationships, only five (14.29%) showed agreement between participant or clinician perception and what was found by either network. Overall, 17 (48.57%) of the relationships noted were perceived by both the participant and clinician but not included in either network. For participants 102, 103 and 104 no cognitive-affect relationships were included in either the estimated contemporaneous or directed network, although relationships between variables were perceived by the participant and/or clinician.

In conclusion, our findings in relation to hypothesis two were that we were unable to find sufficient data in support of the hypothesis. Four 'lead-lag' relationships were included in the directed network for participant 105, but, as no similar relationships were included for other participants, this was not adequate evidence and the hypothesis could not be supported.

However, our further analysis identified three relationships between cognitive and affect variables in the contemporaneous network for participant 101, as well as a clear difference between the number of cognitive-affect variable relationships perceived to exist by both participants and clinicians and yet not included in either network. These additional findings raised the possibility that these findings were produced as a result of the methodology used to model relationships within the contemporaneous and directed networks being, as yet, insufficiently sensitive to detect all relevant relationships. If this is the case, these findings cannot be considered conclusive in terms of the operationalisation of the cognitive-affect variable relationships within the cognitive model of PTSD, although they have implications for our understanding of the limitations of the network models we used in terms of capturing symptom relationships fully.

Discussion

The network approach to psychopathology allows researchers to use data gathered by an individual during the course of their daily life to verify relationships between symptoms which are hypothesized to be maintaining their mental health difficulties. In this study we aimed to use this approach to explore relationships between cognitive and affective symptoms within PTSD, an interaction considered key to symptom maintenance according to the cognitive model of PTSD. We also compared the understanding of a participant's experience of PTSD gained from network analysis with that obtained through clinical formulation, an existing approach to developing this kind of insight. The hypotheses considered were:

Hypothesis 1: Network analysis-derived formulations will resemble therapist-derived formulations in terms of complexity and relationship between elements of the problem.

Hypothesis 2: There will be a 'lead-lag' relationship between changes in cognitive and affect variables.

Our findings in relation to our first hypothesis were that the network analysis-derived formulations did not resemble the therapist-derived formulations in terms of complexity nor in terms of the relationships between elements of the problem. Network-derived formulations included fewer formulation elements, a narrower range of information and were less integrated, so provided a less complete explanation of symptoms than clinician-derived formulations. No clear pattern in terms of the number of relationships was identified, although this may have been partly due to imprecise regularization leading to the inclusion of false positive relationships within one participant's directed network. Relationships within the two approaches overlapped significantly in terms of many symptom connections, but differed in the important respect that all relationships within the network analysis-derived formulations were idiographic and based on that participant's data, whereas relationships between pre-determined elements of the clinicians' formulations were not clearly established as necessarily true of that participant, rather than being dictated by the cognitive model. This meant that the only clearly idiographic relationships within the clinicians'

formulations were those adding personalised information to the formulation. Relationships within network analysis-derived formulations were also limited to showing co-occurrence of symptoms within a window of measurement (contemporaneous networks) or a predictive relationship between them (directed networks) which meant they failed to include relationships between symptoms which did not fit these criteria, excluding in some cases the influence of symptoms rated as important by the participant and noted as central to the clinicians' formulations.

Our second hypothesis regarding a lead-lag relationship between changes in cognitive and affective variables was also not upheld. Although we found four direct relationships between cognitive and affective variables in the estimated directed networks, this was insufficient evidence to support the hypothesis. Evaluation of the contemporaneous networks, in which we found further relationships between cognitive and affective variables, as well as examination of perceived relationships between cognitive and affective variables, which indicated that cognitive-affective variable relationships were perceived to exist more frequently than had been estimated in the networks, suggested that these findings may reveal the limitations of our estimated networks rather than being an accurate reflection of symptom interactions.

A number of factors limited our findings for the first hypothesis. While both the clinician-derived and network analysis-derived formulations were based on the cognitive model of PTSD, and so included pre-determined elements taken from that model, overall the network-derived formulation was more constrained than the clinician-derived formulation. Network formulations were limited in their explanatory power not only by the number and type of variables it was possible to include, but also by those variables being limited to choices from standardised scales. These limits arose primarily out of concern to ensure valid measurement of constructs, which implied the use of properly validated variables such as those from standardised scales, and the use of a minimum of three variables to maximise the likelihood of properly capturing each different parts of the formulation, for example, the inclusion of three separate cognitions. However, the combination of these requirements with limits on the total number of variables it was possible to include, due to the

need to manage the burden on respondents, resulted in the network-derived formulations being both less idiographic and less comprehensive in their explanation of symptoms than the clinician-derived formulations.

It is important to note that variability in our results, for example, in the number of elements included within clinician-derived formulations, and specifically the bi-modal distribution found when means for the number of elements included in the formulations for participants 101-103 and 104-105 were calculated separately, indicate that individual clinician factors were also likely to have influenced our findings. Using five different therapists to produce the clinician-derived formulations meant that individual clinician differences were not controlled for, and it was not possible to identify potential confounding variables. The variation in clinical experience between therapists participating in the study may have been one possible influence on findings, although it not the case that formulations 101-103, which included the greatest number of elements, were produced by the three most experienced therapists. Therefore, other individual factors relating to the clinicians may also have been important. Whatever these factors were, and while, in terms of our analysis of the number of formulation elements included, the variability between formulations did not change our overall conclusion, it is clear that these findings must be viewed with some caution given the likely influence of individual factors on aspects of formulation under consideration, as well as the small sample size.

Some of the differences found between the two formulation approaches may be reduced by maximising personalisation through the use of idiographic, rather than standardised variables, which appears justifiable given the use of wholly idiographic information in comparable clinician-derived formulations. If this were to be done, care would have to be taken to ensure that variables mapped independent factors rather than aspects of the same latent construct, which could lead to misleading correlations between variables which were in fact just different descriptions of the same symptom (Birkeland et al., 2020). Variables could also be selected which reflected inferred information to a greater degree, potentially by combining clinician and participant perspectives on

the variables to include, as proposed in previous studies (Epskamp et al., 2018; Piccirillo & Rodebaugh, 2019). However, some differences identified by our research appeared to arise from fundamental dissimilarities in the approaches which may be more difficult to resolve. The scope of network-derived formulations is likely always to be more focused than clinician-derived formulations, due to restriction, as noted, on the number of included variables arising from natural limits due to the burden of data collection. In terms of the relationships identified, if the failure of network-derived formulations to include the influence of variables which do not co-vary with others (for example, difficulty with anger management for participant 103) is due to ceiling effects in the rating scales used, this may be overcome by adjusting the relevant scales. However, if lack of co-variance within a chosen window of measurement between a variable of interest and other variables included occurs because it is more stable than other variables, or because its effects unfold on a different time scale, an issue with network-derived formulations previously identified (Robinaugh et al., 2020), then any network whose edges are estimated based on co-occurrence of symptoms is unlikely to capture this variable's influence, even when, as for participant 103, it is rated as significant by both the participant and clinician.

In contrast, however, network analysis-derived formulation offered incremental validity over clinical formulation in one respect, by including only empirically based idiographic inter-variable relationships, rather than a mix of idiographic and pre-determined relationships which had not been validated for that particular participant. This difference could be applied within clinical settings to identify symptom interactions precisely. Previous studies have suggested that network analysis and clinical formulation could provide complementary insights into psychological difficulties, for example in creating a personalised intervention plan based on the symptoms found to be most important to a client struggling with depression and social anxiety (David et al., 2018). Our findings offer support for this case, despite a previous comparative study (Schumacher et al., 2021) which contrasted estimated networks of trauma symptoms based on clinicians' perceptions and empirical data from refugee youth participants, and interpreted a weak association between the structures of the two

networks as indicating potential lesser clinical validity of the empirical network, despite good agreement of more than 60% on the presence/absence of edges. While we also found differences between clinical and network-derived formulations, these appeared more associated with methodological strengths and limitations of the two approaches than attributable specifically to a lack of clinical validity in the network analysis. Overall, therefore, our findings supported the potential clinical use of network approaches, in agreement with the argument that techniques such as network analysis, which aim to formalise case conceptualisation, help to bridge the gap between clinical practice and mental health research and foster more scientific practices in tailoring treatment by requiring clinicians to be more rigorous, especially in specifying relations within their formulations (Burger et al., 2020).

The fact that we did not obtain sufficient support for hypothesis two, finding only four examples of a cognitive variable predicting an affective response variable, or vice versa, in the estimated directed networks, may be attributable to the use of insufficiently idiographic variables. Earlier researchers noted that "any systematic comparison between theory-implied and empirical data models would require that variables used in data collection either directly map onto components in the theory, or that they can be precisely derived from those components" (Burger et al., 2020, p. 12). By using variables derived from standardised measures we are likely to have limited the ability of our study to capture the most salient thoughts and feelings for each of our participants, thereby reducing the capacity of our estimated network to include the key interactions hypothesised to drive their symptoms, according to the cognitive model of PTSD (Ehlers & Clark, 2000). As a result our study was unable to contribute to understanding of this specific interaction, although it did function as a proof-of-principle study in terms of the potential of network analysis to elucidate these kinds of interactions.

In this study we had attempted to build on the findings of previous research, which established that network analysis could provide insight into the operation of cognitions and emotions within PTSD. A systematic review of studies using the network approach to PTSD

(Birkeland et al., 2020) confirmed the importance of cognitions and emotions, identifying 'recurrent thoughts of trauma' and 'persistent negative emotional state' as repeatedly among the top three most central symptoms within the studies analysed. Greene and colleagues' (2018) network study of PTSD symptoms identified a feedback loop in the directed network between negative beliefs, which predicted avoidance of thoughts, which then predicted negative emotions. A follow-up study based on the same data (Greene et al., 2020) investigated temporal relationships further, testing the hypothesis that negative emotions predicted PTSD symptom clusters but finding the reverse, that PTSD symptom clusters predicted negative emotions. Limits on the number of included variables precluded analysis of negative cognitions as an independent variable in this study, and instead they were grouped with emotions in the PTSD symptom cluster 'negative alterations in cognitions and mood' (NACM). Within the directed network, the NACM cluster was predicted by fear and sadness and predicted the negative emotions sadness and loneliness. As both studies by Greene and colleagues were conducted peri-traumatically with participants who did not all have diagnoses of PTSD, our study was potentially more relevant to developing understanding of cognitive-affective symptom interactions within a clinical population. Although our limited findings were insufficient to support our hypothesis, the study had value in extending insight into how network analysis might be applied to uncover symptom interactions for these participants, with the shortcomings we identified in terms of the required precision of variables an area for consideration in future network analyses.

Furthermore, our findings suggested that achieving the exactness of fit between symptoms and variables necessary for network analysis to offer a better operationalisation of cognitive theory would require not only more precisely tailored variables, but also a more appropriate time lag in the directed network. The additional relationships between cognitive-affective variable pairs found in the contemporaneous network, as well as the high number of relationships between cognitive and affective variables perceived by clinicians and participants but not found in either network, suggest that the time lag chosen, as well as the variables used, may not have been optimised to retrieve all meaningful relationships. The selection of a time lag is difficult, especially when the speed of

variable interaction is unknown. Wichers et al. (2015) claimed that micro-level interactions such as those between affective states have an important role in mental health difficulties, and noted the highly labile nature of variables such as emotions and cognitions, which may make capturing these interactions difficult. Studies using quite different time intervals (90 minutes in Wichers et al., 2014; 12-15 hours in Greene et al., 2018) have succeeded in capturing symptom interactions at the level of emotions, providing evidence that this type of study is possible. However, our findings, together with these earlier studies, imply that it remains challenging to determine appropriate time intervals within network analysis, and we have been unable to find evidence so far of any general rule to apply to this, other than assessment of the idiographic presentation of the individual taking part, and consideration of the purpose of the research.

Limitations

Limitations of this study included first, the small sample of participants. While each participant completed a minimum of 59 data points and their individual results were considered sufficiently robust for analysis, the small number of participants means that findings cannot be extrapolated more widely. Second, our sample was highly likely to be subject to selection bias as it was limited to those who were both diagnosed with PTSD and able to complete data collection multiple times daily. The demands of data collection prevented some potential participants from taking part, indicating that this study, and potentially others with similarly demanding recruitment criteria, are likely to be biased in favour of the select group of digitally literate individuals who are not significantly debilitated by their symptoms and able to tolerate reporting on them multiple times daily. The reliance on participants to choose variables for inclusion may have led to less aversive variables being included, as well as variables into which the participants had insight, which may have left out relevant but upsetting symptoms. Our reliance on participants to complete data collection independently, despite discussion of variables and guidance regarding how to interpret them, may still have led to idiosyncratic interpretation of questions 'in the moment' (as noted in Chun, 2016). Our network analysis assumed stationarity of data, and while all our participants were mid-way

through therapy and data collection took place over a short period of two weeks, we cannot assume that symptoms were consistent over this period. In fact, participant 105 noted that they found the process of frequent reflection on their symptoms rather therapeutic and it was noted that their PCL-5 score declined from 66 to 47 over the course of data collection, while still remaining above cut-off for diagnosis of PTSD. Each participant's data also included a small proportion of missing data, which is typical in network studies. We made the assumption that all missing data was missing completely at random (MCAR) which was not confirmed, and may have had some effect on findings if incorrect.

A particular limitation of the findings in relation to hypothesis one was that every participant in the study had a different therapist for pragmatic reasons. This meant that the influence on formulations provided of individual differences between those therapists, including their differing levels of experience, was not controlled. Research has indicated that therapist experience, in particular, affects formulation, with more experienced therapists having been found to produce formulations that are more parsimonious, internally consistent and contained fewer errors (Dudley et al., 2015). While these were not the specific attributes examined in this study, and our comparison was between clinician-derived and network-derived formulations as a whole, rather than between individual clinician-derived formulations, the potential influence of this factor on the complexity and relationships included within the clinicians' formulations, and so the results of this comparison, should be acknowledged.

This study had a number of strengths in being a proof of principle study which piloted a method of investigating dynamic symptom interactions, as well as a test of an approach to comparing clinician-derived and network-derived formulations. It was a pragmatic study, using data obtained from participants across multiple NHS Trusts, who were receiving treatment for PTSD in real clinical settings throughout the course of the research, which gave the data high ecological validity. As a novel study, the method of synthesizing results had to be developed by the researcher and supervisor, with findings based largely on the researcher's interpretation. Efforts were made to base this synthesis on clear and transparently described criteria, with term definitions taken from

relevant research literature. The process was also clearly described in order to allow replicability. A further strength of this approach was that it could be tailored specifically to the requirements of our study. However, the rather subjective nature of some elements of the analysis, especially the analysis of complexity, does raise questions in relation to study objectivity and replicability. It is recommended that future studies are conducted to verify this study's findings, with more than one researcher producing comparisons of formulation approaches, the degree of correspondence between those comparisons determined and the different findings synthesized, in order to address these issues. In this way, future studies' findings may be more conclusively established to be objective and robust.

Implications

The findings of this study have multiple clinical and scientific implications. Our findings imply that network-derived formulations have potential value within clinical settings as an additional source of insight into a client's difficulties. Network analysis was found to offer specific insight into the idiographic relationships between a client's particular symptoms using a client's own empirical data to document these interactions, information which could not be obtained through clinical formulation. The utility of network analysis-derived formulations' basis in empirical data has been demonstrated previously (Kroeze et al., 2017) when evidence from network analysis of an individual client's difficulties was employed in clinical practice to explore their symptom interactions with the client and inform discussion about suitable treatment. However, previous studies have also noted that the findings of network analysis are not sufficient in themselves to guide decisions about treatment interventions (Greene et al., 2020). Our study supported this conclusion by confirming network analysis-derived formulations' relative sparseness and lack of explanatory power compared with clinicians' formulations, suggesting that it may be better used for focused exploration of specific hypotheses originated in clinical formulations, as has also been proposed by previous researchers in network analysis (David et al., 2018; Robinaugh et al., 2019). Questions remain about the appropriate clinical application of network analysis, including the most appropriate stage of

treatment at which to consider it and the reasonable duration of data collection during treatment (Epskamp et al., 2018); although a two-week data collection mid-treatment was reasonably well tolerated by our participants, they constitute too small a sample from which to draw firm conclusions in respect of these questions.

The scientific implications of this study relate to both hypotheses tested. First, despite the limited findings of our network analysis of interactions between cognitive and affective variables, this study confirmed that network analysis is an appropriate method through which to investigate these relationships; it functioned as a successful proof-of-principle in terms of using this approach. The limitations of what we found have implications for the conduct of future network studies evaluating these kinds of symptom-level interactions.

Network researchers have acknowledged the challenge of identifying the most appropriate intervals of measurement for network analyses (Epskamp et al., 2018) and this was also a question raised by our evaluation of interactions between cognitive and affective variables. Although earlier studies (Greene et al., 2018; Greene et al., 2020) had documented symptom-level interactions using a 12-16 hour window of measurement, we were able to identify only a very small number of cognitive-affective variable interactions within our directed network using a three hour window, although more were found in the contemporaneous network. Differences between the studies in terms of participants, contexts and variables used may explain these different findings in part, in addition to the different window of measurement. Nonetheless our limited findings draw attention again to the question of how best to identify appropriate intervals of measurements for network studies, where this key decision is required be made in advance of data collection and is also determined by practical considerations such as limitations arising from not creating an unmanageable burden of data collection. While our findings, together with those of previous studies, suggest that these decisions may be best determined on a case-by-case basis, questions remain over whether any particular principles or considerations may still be useful in guiding this assessment.

The implication of our conclusion that the network analysis of cognitive-affective variable interactions may have been limited by our use of standardised variables is that researchers undertaking these types of studies should prioritise ensuring that variables measured are the most salient possible for their participants over using pre-existing validated measures. While we wanted to ensure that the variables we used had been established as valid, this appears to have inhibited the capacity of our study to examine the central interactions we wished to study.

In terms of the comparison of network analysis-derived and clinician-derived formulations, this study confirmed previous observations that the assumptions made when both contemporaneous and directed networks are estimated risk leaving out important aspects of a participant's presentation (Robinaugh et al., 2020). Examples of these assumptions include the idea that variables within a psychopathological network vary in a way that is clearly positive or negative and symmetrical (for example, that they clearly exacerbate or inhibit one another, rather than have a slight moderating effect, and that this effect is equal in both directions) or that symptoms present in a uniform way, rather than being dimensional (although Borsboom (2008) has challenged the idea of dimensional symptoms). In this study contemporaneous networks were estimated by analysing co-variance within our window of measurement, which was three hours. As a result, our networks reflected the pattern of relationships found under this condition but potentially excluded relationships which unfolded over a longer time frame. Further consideration of how to reflect the contribution of variables whose effects may unfold over a different time scale to other relevant variables, but whose exclusion from networks weakens those networks capacity to paint a complete picture of symptom interactions is recommended.

Conclusion

This study considered two hypotheses; that network-analysis derived and clinician-derived formulations resembled one another in terms of complexity and relationships between elements of the problem and whether a 'lead-lag' relationship existed between cognitive and affect variables for participants diagnosed with PTSD. Our findings did not uphold either hypothesis, although the study

demonstrated that network analysis was a reasonable approach through which to consider these questions as well as identifying ways in which future studies may refine the approach used in order to achieve more definitive results.

Further, the study established that network analysis had potential to offer useful clinical information not possible to obtain through clinical formulation, specifically by using participant's own empirical data to confirm the idiographic relationships between their symptoms. Our study also suggested ways in which clinical insight may be used to develop network analysis by guiding definition of the idiographic symptoms to include.

Finally, the study clarified some limitations of network analysis that may benefit from further theoretical consideration. In particular, the study highlighted the question of how to incorporate within networks those symptoms which do not vary in tandem with other symptoms, meaning their relationships and influence are hard to evaluate within the constraints of current network analytical protocols.

Further Research Recommendations

Research to develop the findings of this study should include a follow-up study to evaluate cognitive-affective variable relationships within PTSD using more personalised variables, in order to further evaluate this aspect of the cognitive model of PTSD.

We observed that self-blame, shame and guilt were often symptoms noted by clinicians to maintain client difficulties, but not included in our network analyses. Blame of self or others was noted by Greene et al. (2018) as having a mediating role in the relationship between negative emotions and negative beliefs. A network analysis considering interactions between blame and these variables for participants who experience this may provide useful insights into the potential contribution of blame to PTSD symptom maintenance.

Finally, our findings that network analysis could be applied clinically as a means of validating hypothesized symptom relationships should be further tested. Wichers et al. (2017) has recommended that randomised controlled trials to evaluate whether insights into a patient's

network dynamics can significantly improve clinical decision making and patient outcomes should be commissioned, and we support this recommendation. Given the burden of data collection on participants, establishing that network analysis can offer sufficient benefit to clients in relation to the effort required is an important consideration in understanding its potential for use in clinical practice.

Integration, Impact and Dissemination

Integration, Impact and Dissemination

Integration

The overall objective of this thesis was to contribute to current thinking about what a good understanding of an individual's psychological difficulties should be like and how this might be achieved. The central question bringing the two studies together could be framed as 'How best can we capture and measure an individual's difficulties in a formulation that both includes the richness of their experience and is justifiable in empirical terms?' Both the systematic review and empirical study examined methods of developing this kind of understanding and evaluated them critically. The systematic review considered criteria used to assess formulation, focusing principally on the use of reliability and validity as measures by which to evaluate cognitive formulations, examining the adequacy of these and considering the degree to which quality assessment might improve on or supplement them. The empirical study compared the relatively new approach of network analysis-derived formulation with cognitive formulation for post-traumatic stress disorder (PTSD), again in order to understand their relative merits and limits. The empirical study also examined whether it was possible to identify cognitive-affective interactions through network analysis, in particular the cognitive-affective interactions considered central to the cognitive model of PTSD; a question not only interesting in its own terms, but which also shed light on the capabilities and limits of this approach in terms of understanding individual psychopathology. The systematic review and empirical paper were not dependent on one another's conclusions but rather linked thematically.

As a core competency for clinical psychologists, formulation is a central part of clinical practice (HCPC, 2015). Psychologists are required to evaluate their practice systematically (HCPC, 2015) and yet formulation evaluation is complicated by a lack of clarity regarding suitable evaluation criteria. While formulation is understood to combine knowledge of scientific research with clinical expertise tailoring this knowledge to a particular client, formulations nonetheless rely on individual psychologists' interpretation of the source information. Acknowledging this potential for subjectivity and error, within cognitive therapy it is stressed that formulations are hypotheses, and should be

viewed as tentative (Leahy & Beck, 1996; Persons, 1989). Even so, it is perhaps unsurprising that within the context of the 'scientist-practitioner' model of psychology, criteria such as reliability and validity, which are usually applied to psychometric measures and link to notions of scientific objectivity and integrity, have been applied to the assessment of formulation. The use of these criteria suggests a concern that the idiographic nature of formulation means it is somehow insufficiently 'scientific' and empirically grounded, an anxiety which researchers have tried to address by confirming that formulations meet these standards.

My review of studies evaluating the degree to which formulation could be considered reliable and valid highlighted how far these concepts were helpful in terms of assessing formulation. Overall, reliability was found not to be an especially useful criterion for evaluating formulation. Two formulations could, quite reasonably, differ when they focused on different aspects of a client's presentation, and yet both be helpful in determining appropriate treatment. Attempts to assess validity were varied in approach although validity was frequently evaluated using indirect measures only, which complicated assessment of its utility. However, some studies, especially those of Mumma and colleagues (Mumma, 2004; Mumma & Mooney, 2007; Mumma & Fluck, 2016) successfully assessed validity directly using a method which was both collaborative between client and clinician and also made use of the client's own personal data to confirm the hypotheses tested. This provided helpful insights however, as a gauge of how well a formulation meets its aims, validity was less comprehensive than measures of formulation quality also evaluated in the systematic review (see Eells et al., 1998; Kuyken et al., 2005; Haarhoff et al., 2011). These tools attempted to draw together knowledge about formulation to present an objective view of its necessary components, not only in terms of content categories, but also qualitative aspects, for example, complexity, degree of inference and precision of language (Eells et al., 1998). While the exact criteria included varied between different quality assessment tools, both content and qualitative aspects of formulation were included in every quality assessment measure in this review. This suggested a degree of consensus regarding the importance to a 'good' formulation both of observed symptoms,

and the degree to which that formulation synthesized those observations to explain and give meaning to them. While the development of these tools is an alternative to the psychometric model of formulation evaluation represented by reliability and validity assessment, their core aim remained to produce objective standards against which formulation can be usefully evaluated.

Rather than examining assessment criteria, my empirical study focused on methods, comparing the kind of understanding of the difficulties experienced by someone with PTSD produced by a network-analysis derived formulation with that of a therapist-derived formulation. Given that a network analysis-derived formulation is driven by personalised data collected from the participant, this comparison may to some extent also be said to have operationalised questions brought up by the systematic review concerning whether a potentially less subjective, data-driven approach to formulation might produce better or different insight to a clinician. The comparison suggested that overall, clinician-derived formulations provided a broader range of information and synthesized this more effectively into an explanation of the participant's difficulties than the network-derived formulation. However, the finding that the network-derived formulation was more precise and idiographic in terms of the relationships between elements of the formulation depicted, suggested that being founded on data allowed it to distinguish those relationships which were really relevant to that individual, rather than include all possible connections, despite the fact that both approaches to formulation were based on the same cognitive model of PTSD. While limitations in terms of the scope of network analysis-derived formulations compared to clinician-derived formulations may be difficult to overcome, our findings suggest that using more personalised variables had potential to improve the value of the insights offered by network analysis.

The final part of the study, which was the analysis of whether a 'lead-lag' relationship could be identified within the contemporaneous and temporal networks produced following data collection from clients with a diagnosis of PTSD, stood slightly separately to the other parts of this thesis. However, it connected thematically in helping demarcate what was possible in terms of understanding symptom interactions within this type of approach. The capacity of network analysis

not only to validate relationships between symptoms, but also to identify when one symptom predicts another, expands the scope of information available to clinicians in developing their understanding of client difficulties.

In terms of the larger theme of this thesis, the findings of the empirical study made the case that both clinician-derived formulations and network-derived formulations have different qualities, all of which may contribute to a good understanding of difficulties. This raises the question of whether any single approach to formulation can capture all of the information which may be helpful and relevant to clinical practice. In conclusion, the formulation types may be viewed as complementary, operating at different levels of analysis (Persons & Tompkins, 2007). How far a formulation is perceived to be 'good' is likely to depend at least in part on its parameters and purpose, suggesting formulation evaluation should reflect those dimensions. Network analysis may not be the best method with which to draw out the meaning of symptoms, but could be used to direct treatment in different ways, such as identifying precisely a developing dysfunctional interaction.

Reflections on Research Process

Systematic Review

The process of completing this thesis has not been smooth. Due to a combination of difficult personal circumstances and the advent of a pandemic I had to extend my doctorate and change the topic of my empirical study several times, both of which affected this thesis.

Uncertainty over the final topic of my empirical paper, which changed following me having to extend my doctorate, contributed to the decision to focus on reliability and validity; it was felt that there was sufficient research into these concepts in relation to formulation to justify a systematic review, and that this review would complement any of the options considered for my empirical paper. However, while conducting scoping searches, I identified recently conducted systematic reviews which covered very similar ground to my original plan (Easden & Kazantzis, 2018; Flinn et al., 2015). What was clear after reading these reviews was that, while they contained useful insights, they left

room for further interrogation of the definitions of reliability and validity used in the studies reviewed. I felt there was also the opportunity to include reflection on developments in assessment of formulation quality, which I perceived to be highly relevant as a further attempt to apply objective, standardised criteria to formulation. What was challenging when I conducted the review was that there was insufficient reporting of raw formulations within the clinical papers to allow independent evaluation of formulation quality and so testing of the quality criteria for formulation we had devised. This limited the conclusions I was able to draw regarding the degree to which these criteria might be useful. Nonetheless the completion of the systematic review was an interesting exercise in terms of thinking through criteria for formulation evaluation and also led to further reflections.

Assessing the ways in which reliability and validity were defined in the context of conducting the systematic review, and discovering the degree to which these definitions were often partisan, also allowed me insight into the ways in which all researchers, no matter how much we strive for objectivity, are prey to our own biases and personal interests. While this is understandable, it led me to consider again the effect this may have on research, and the need consistently to interrogate the terms in which research is defined and conducted whenever I review or conduct research in future.

Empirical Study

While my empirical study was always intended to be small in scale, I encountered significant difficulties with recruitment which meant that by my original submission deadline I had managed to recruit only two participants after having approached seven NHS services, and gained NHS ethical approval for four. This necessitated a six month extension. It had been clear that the decision to continue to pursue a project involving NHS services and clinicians in the middle of the Covid-19 pandemic, during which services were often reduced in scope and under extreme pressure, as well as seeking participants with active symptoms of PTSD to carry out multiple surveys daily, meant that recruitment would not be easy. The resulting small sample may have limited the findings of our study, given that lack of power is understood to reduce the capacity to identify correlations within

networks (Epskamp, 2018). In order to be able to compute networks we required a minimum number of data points which meant we asked participants to complete measures five times a day for two weeks. I had piloted the study personally and found data collection intense. The demands of this schedule on clients diagnosed with PTSD in the severe range was significant and may have contributed to difficulties with recruitment as well as missing or inaccurate data, affecting our findings. My reflection, which arose from this experience, that it was important to balance the demands of network analysis with better understanding of its benefits to clients in terms of clinical outcomes was included in the further research section of the empirical study.

Impact

Limitations identified in both the systematic review and empirical study mean that the conclusions drawn from this thesis are tentative, but nonetheless potentially useful in shaping the approach of future research and clinical practice.

Impact on Future Research

Systematic Review. The systematic review highlighted an important concern for clinical psychologists, which is the lack of appropriate, objective criteria by which to evaluate their formulations. Given the requirement to endeavour that practice is empirically-based (HCPC, 2015) it is hoped that this study will direct attention to the need to develop these criteria. The review found that, although there has been significant research focus on evaluation of reliability and validity as criteria for formulation evaluation (e.g. Kuyken et al., 2005; Mumma & Fluck, 2016), quality assessment may represent a more comprehensive and objective standard. This adds weight to arguments for further research into formulation quality, specifically into whether criteria included in existing tools of quality assessment such as the CFCCM (Eells et al., 1998) or the quality assessment tool used in our study have clinical utility. Further research in this area may also be guided by the insight from this review that previous studies adopted 'expert' views as measures of formulation reliability (e.g. Haarhoff et al., 2011), contrary to the ethos of cognitive formulation which emphasizes collaboration. This confirms the need for future research in this area to scrutinize the

objectivity of assessment criteria. Finally, it is hoped that by contributing to understanding of ways in which formulation may be evaluated, the review may thereby contribute, albeit in a limited way, to the larger aim of understanding the aspects of 'good' formulations most critical to making them clinically useful.

Empirical Study. As far as we could determine, the empirical paper was the first study to estimate contemporaneous and temporal networks for participants with a confirmed diagnosis of PTSD. As such, the networks produced are of interest to researchers engaged in the fields of both network analysis and PTSD and it is hoped will encourage similar studies. A couple of specific areas are especially noteworthy in terms of impact on future research. First, the challenges of capturing fast-changing internal phenomena such as emotions using temporal networks noted in previous research (Greene et al., 2020) was confirmed. Previous studies (Wichers, 2014) have used time-sampling as frequent as every 90 minutes to capture changes in variables relating to emotional state, while others have captured variable interactions using a window of measurement as long as 12-15 hours (Greene et al., 2018). Our study used a three hour interval and succeeded in capturing a limited number of temporal relationships between cognitive and affect variables, although too few to draw robust conclusions. Our finding in terms of the effect of using variables in the network analysis drawn from standardised measures, rather using idiographic variables, is also useful for future studies, suggesting that using personalised variables is important to capturing key symptom interactions. While our attempt to operationalise cognitive theory regarding cognitive-affective variable interactions within PTSD was of limited success, the study functioned as a proof-of-principle and it is hoped that future research will be able to utilise our findings to produce a more successful and sufficiently powered operationalisation of this theory using network analysis, as proposed in Robinaugh (2020). Burger et al. (2020) also emphasized the need for research to formalize idiographic theories, arguing that research using the opportunities inherent in the network approach to develop broader theoretical understanding had advantages for both clinical practice and mental health research, potentially bridging the scientist-practitioner gap.

Community and Clinical Impact

Systematic Review. It is hoped that the findings of my systematic review in relation to the potential value of quality assessment of formulation may have clinical utility for psychologists who wish to evaluate their formulations. While more research is needed to confirm the most appropriate criteria within quality assessment, the systematic review may also be useful in drawing attention to the need for clinicians using formulation to be accountable for the methods they use and ensure they meet appropriate standards. Given the importance of formulation within clinical psychology previously noted, it is hoped that these findings will encourage greater scrutiny of formulation practice, as well as offering means by which formulation can be assessed, thereby ensuring formulations used within clinical practice are of a standard most likely to benefit clients.

Empirical Study. One important community impact of the empirical project may be on the participants of this study, as I intend to share their network maps with them and their clinicians. It is hoped that this information may offer helpful insight as they continue in therapy, as has been found in previous research in which feedback from network analysis has been shared (Kroeze et al., 2017). Both clients' and clinicians' reflections on this experience will be monitored.

Participant 105 shared the reflection that they had found the data collection process helpful, as they had experienced the symptom monitoring as something that allowed them to reflect on and vent their feelings. Examination of this participant's PCL-5 scores showed a notable decrease over the course of data collection (from 66 to 47), although they still met diagnostic criteria for PTSD at the end of data collection. While numerous factors may have contributed to this change over the course of this participant's engagement in this study, and this sample consisted of a single participant, it may be useful in future studies to monitor whether there is a subset of people who receive a therapeutic effect from symptom monitoring conducted even in this intense way, for consideration as part of the clinical application of network analysis.

This research is still in its infancy in terms of contributing to theoretical understanding of PTSD. However, as an early attempt to map symptom interactions over time within PTSD, it is hoped

that this project will contribute to developing understating of these relationships and that that knowledge will be of benefit not only to research but also to the treatment of PTSD by increasing the range of tools available to clinicians to understand their clients' experiences and offering clients and clinicians the opportunity to develop network analysis-driven formulations of their symptoms which draw on their own idiographic data.

Dissemination

Dissemination of research findings has an important role in communicating the outcomes of research to different audiences who may find them helpful and ensuring that research achieves the greatest possible impact.

Initial findings from the empirical study were disseminated via a presentation to staff and current D.Clin.Psy. students at Royal Holloway, University of London in May 2021. The aim of this presentation was primarily to explain the rationale and methodology of this study, helping develop fellow professionals' understanding of network analysis and thereby encouraging further studies in this area.

Findings of the empirical study will also be communicated to clinicians who took part in the study in order to share learning and encourage reflections on the process, which I hope may prove helpful to developing the methodology used for future research projects.

To maximise the potential impact of this research, following consultation with the thesis supervisors, I hope to submit either or both the systematic review and empirical study to appropriate academic journals for peer review and eventual publication. A number of recent studies employing similar methodology (Greene et al., 2020) have been published in the Journal of Traumatic Stress, which is an established publication within the field of research into PTSD, with influence among both clinicians and researchers. An alternative would be the journal Psychological Medicine, which has also published studies in this area and which is also a well-known and reputable publication. Publishing the study findings in either of these journals would assist the dissemination of our findings significantly.

In addition, I hope to present the findings of the empirical study to at least one conference attended by researchers engaged with network analysis studies and/or its application to PTSD. Appropriate conferences at which I hope to offer a paper or poster are those organised by the UK Psychological Trauma Society or the European Society for Traumatic Stress Studies, or, in the field of network analysis, the winter or summer schools organised by the Psychosystems project team.

References

Abel, A., Hayes, A. M., Henley, W., & Kuyken, W. (2016). Sudden gains in cognitive–behavior therapy for treatment-resistant depression: Processes of change. *Journal of Consulting and Clinical Psychology*, *84*(8), 726–737. https://doi.org/10.1037/ccp0000101

Addis, M. E., & Jacobson, N. S. (2000). A closer look at the treatment rationale and homework compliance in cognitive-behavioral therapy for depression. *Cognitive Therapy and Research*, *24*(3), 313–326. https://doi.org/10.1023/A:1005563304265

Alexander, L. B., & Luborsky, L. (1986). The Penn Helping Alliance Scales. In L. S. Greenberg, & W. Pinsof (Eds.), *The psychotherapeutic process: A research handbook*, 325-366. The Guilford Press.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Bak, M., Drukker, M., Hasmi, L., & Os, J. van. (2016). An n=1 clinical network analysis of symptoms and treatment in psychosis. *PLOS ONE*, *11*(9), e0162811. https://doi.org/10.1371/journal.pone.0162811

Barnow, S., Stopsack, M., Grabe, H. J., Meinke, C., Spitzer, C., Kronmüller, K., & Sieswerda, S. (2009). Interpersonal evaluation bias in borderline personality disorder. *Behaviour Research and Therapy*, *47*(5), 359–365. https://doi.org/10.1016/j.brat.2009.02.003

Barris, B. P. (1996). *Developing reliable case formulations and individualized treatment plans for a depressed inpatient population* (Doctoral thesis, University of Memphis, USA). ProQuest Information & Learning. http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1996-95023-166&site=ehost-live

Beck, A.T. (1976). *Cognitive therapy and the emotional disorders*. International Universities Press.

Beck, A.T., Rush, A.J., Shaw, B.F., & Emery, G. (1979). *Cognitive therapy of depression*. The Guilford Press.

Beck, A.T., Steer, R.A., & Brown, G.K. (1996). *Manual for the Beck Depression Inventory (2nd ed.).* The Psychological Corporation.

Beck, J. S. (1995). *Cognitive therapy: basics and beyond.* The Guilford Press.

Bieling, P. J., & Kuyken, W. (2003). Is cognitive case formulation science or science fiction? *Clinical Psychology: Science and Practice*, *10*(1), 52–69. https://doi.org/10.1093/clipsy.10.1.52

Birkeland, M. S., Greene, T., & Spiller, T. R. (2020). The network approach to posttraumatic stress disorder: A systematic review. *European Journal of Psychotraumatology*, *11*(1), 1700614. https://doi.org/10.1080/20008198.2019.1700614

Birkeland, M. S., & Heir, T. (2017). Making connections: Exploring the centrality of posttraumatic stress symptoms and covariates after a terrorist attack. *European Journal of Psychotraumatology*, *8*(sup3), 1333387. https://doi.org/10.1080/20008198.2017.1333387

Blevins, C.A., Weathers, F.W., Davis, M.T., Witte, T.K., Domino, J.L. (2015). The posttraumatic stress disorder checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress, 28(6)*, 489-498. https://doi.org/10.1002/jts.22059

Boorsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology, 64(9)*, 1089-1108. https://doi.org/10.1002/jclp.20503

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

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLOS ONE*, *8*(4), e60188. https://doi.org/10.1371/journal.pone.0060188

British Psychological Society. (2017). *Practice guidelines: Third edition.* https://www.bps.org.uk/news-and-policy/practice-guidelines

Bryant, R. A., Creamer, M., O'Donnell, M., Forbes, D., McFarlane, A. C., Silove, D., & Hadzi-Pavlovic, D. (2017). Acute and chronic posttraumatic stress symptoms in the emergence of

posttraumatic stress disorder: A network analysis. *JAMA Psychiatry*, *74*(2), 135–142. https://doi.org/10.1001/jamapsychiatry.2016.3470

Bucci, S., French, L., & Berry, K. (2016). Measures Assessing the Quality of Case Conceptualization: A Systematic Review. *Journal of Clinical Psychology*, *72*(6), 517–533. https://doi.org/10.1002/jclp.22280

Burger, J., van der Veen, D. C., Robinaugh, D. J., Quax, R., Riese, H., Schoevers, R. A., & Epskamp, S. (2020). Bridging the gap between complexity science and clinical practice by formalizing idiographic theories: A computational model of functional analysis. *BMC Medicine*, *18*(1), 99. https://doi.org/10.1186/s12916-020-01558-1

Burns, D.D. (1998). *Therapist toolkit.* [Unpublished manuscript].

Butler, G. (1998). Clinical formulation. In Bellack, A.S. & Hersen, M. (Eds.) *Comprehensive clinical psychology. Volume 6: Adults: Clinical formulation & treatment*, 1-24. Pergamon.

Cahill, J., Stiles, W. B., Barkham, M., Hardy, G. E., Stone, G., Agnew-Davies, R., & Unsworth, G. (2012). Two short forms of the Agnew Relationship Measure: The ARM-5 and ARM-12. *Psychotherapy Research, 22(3)*, 241-255.

Chadwick, P., Williams, C., & Mackenzie, J. (2003). Impact of case formulation in cognitive behaviour therapy for psychosis. *Behaviour Research and Therapy*, *41*(6), 671–680. https://doi.org/10.1016/S0005-7967(02)00033-5

Chadwick, P. D., Lees, S. J., & Birchwood, M. (2000). Beliefs about voices questionnaire—revised (BAVQ-R). *British Journal of Psychiatry, 177*, pp. 229–232.

Chun, C.A. (2016). The expression of posttraumatic stress symptoms in daily life: A review of experience sampling methodology and daily diary studies. *Journal of Psychopathology and Behavioural Assessment, 38*(3), 406-420. https://doi.org/10.1007/s10862-016-9540-3

Clark, D. A., & Beck, A. T. (with Alford, B. A.). (1999). *Scientific foundations of cognitive theory and therapy of depression.* Wiley.

Clohessy, S., & Ehlers, A. (1999). PTSD symptoms, response to intrusive memories and coping in ambulance service workers. *British Journal of Clinical Psychology, 38*(3), 251-65. https://doi.org/10.1348/014466599162836

Cooper, Z., & Fairburn, C. G. (1987). The eating disorders examination: A semi-structured interview for the assessment of the specific psychopathology of eating disorders. *International Journal of Eating Disorders, 6*, 1–8.

Cooper, P. J., Taylor, M. J., Cooper, Z., & Fairburn, C. G. (1987). The development and validation of the Body Shape Questionnaire. *International Journal of Eating Disorders, 6*, 485–494.

Crits-Christoph, P., Cooper, A., & Luborsky, L. (1988). The accuracy of therapists' interpretations and the outcome of dynamic psychotherapy. *Journal of Consulting and Clinical Psychology*, *56*(4), 490–495. https://doi.org/10.1037/0022-006X.56.4.490

David, S.J., Marshall, A., Evanovich, E.K., & Mumma, G.H. (2018). Intraindividual dynamic network analysis – implications for clinical assessment. *Journal of Psychopathology and Behavioral Assessment, 40*, 235.248. https://doi.org/10.1007/s10862-017-9632-8

Davidson, K. M. (2008). Cognitive–behavioural therapy for personality disorders. *Psychiatry*, *7*(3), 117–120. https://doi.org/10.1016/j.mppsy.2008.01.005

Demaria, T. P., Kassinove, H., & Dill, C. A. (1989). Psychometric properties of the survey of personal beliefs: A rational-emotive measure of irrational thinking. *Journal of Personality Assessment*, *53*(2), 329. https://doi.org/10.1207/s15327752jpa5302_10

Derogatis, L. (1983). *The Symptom Checklist-90 Revised: Administration, scoring, and procedures manual II.* Clinical Psychometric Research.

DiGuiseppe, R., Leaf, R., Exner, R., & Robin, M.W. (1988). The development of a measure of irrational/rational thinking. [Paper presentation]. *World Congress on Behavior Therapy, Edinburgh, Scotland, September 1988.*

Division of Clinical Psychology. (2010). *The core purpose and philosophy of the profession*. https://www.bps.org.uk/member-microsites/division-clinical-psychology

Dobson, K.S., & Dozois, D.J.A. (Eds.). (2019). *Handbook of cognitive-behavioral therapies* (4th ed.). The Guilford Press.

Doron-LaMarca, S., Niles, B. L., King, D. W., King, L. A., Pless Kaiser, A., & Lyons, M. J. (2015). Temporal associations among chronic PTSD symptoms in U.S. combat veterans. *Journal of Traumatic Stress*, *28*(5), 410–417. https://doi.org/10.1002/jts.22039

Dudley, R., Ingham, B., Sowerby, K., & Freeston, M. (2015) The utility of case formulation in treatment decision making; the effect of experience and expertise. Journal of Behavior Therapy and Experimental Psychiatry, 48, 66-74. https://doi.org/10.1016/j.jbtep.2015.01.009

Dudley, R., Park, I., James, I., & Dodgson, G. (2010). Rate of agreement between clinicians on the content of a cognitive formulation of delusional beliefs: The effect of qualifications and experience. *Behavioural and Cognitive Psychotherapy*, *38*(2), 185–200. https://doi.org/10.1017/S1352465809990658

Easden, M.H., & Kazantzis, N. (2018). Case conceptualization research in cognitive behavior therapy: A state of the science review. *Journal of Clinical Psychology, 74*, 356-384. https://doi.org/10.1002/jclp.22516

Eells, T.D., Kendjelic, E.M., & Lucas, C.P. (1998). What's in a case formulation?: Development and use of a content coding manual. *Journal of Psychotherapy Practice and Research, 7(2)*, 144-153.

Eells, T. D., Lombart, K. G., Kendjelic, E. M., Turner, L. C., & Lucas, C. P. (2005). The quality of psychotherapy case formulations: A comparison of expert, experienced, and novice cognitive-behavioral and psychodynamic therapists. *Journal of Consulting and Clinical Psychology*, *73*(4), 579–589. https://doi.org/10.1037/0022-006X.73.4.579

Eells, T. D., Lombart, K. G., Salsman, N., Kendjelic, E. M., Schneiderman, C. T., & Lucas, C. P. (2011). Expert reasoning in psychotherapy case formulation. *Psychotherapy Research*, *21*(4), 385–399. https://doi.org/10.1080/10503307.2010.539284

Ehlers, A., & Clark, D.M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy, 38*, 319-345. https://doi.org/10.1016/S0005-7967(99)00123-0

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry*, *33*(6), 766–771. https://doi.org/10.1001/archpsyc.1976.01770060086012

Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A.-M., Riese, H., & Cramer, A. O. J. (2018). Personalized network modelling in psychopathology: The importance of contemporaneous and temporal connections. *Clinical Psychological Science*, *6*(3), 416–427. https://doi.org/10.1177/2167702617744325

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE–OM. *The British Journal of Psychiatry, 180(1)*, 51-60.

Fennell, M. J. V., & Teasdale, J. D. (1987). Cognitive therapy for depression: Individual differences and the process of change. *Cognitive Therapy and Research*, *11*(2), 253–271. https://doi.org/10.1007/BF01183269

Flinn, L., Braham, L., & das Nair, R. (2015). How reliable are case formulations? A systematic literature review. *British Journal of Clinical Psychology*, *54*(3), 266–290. https://doi.org/10.1111/bjc.12073

Flitcroft, A., James, I. A., Freeston, M., & Wood-Mitchell, A. (2007). Determining what is important in a good formulation. *Behavioural and Cognitive Psychotherapy*, *35*(3), 325–333. https://doi.org/10.1017/S135246580600350X

Foa, E. B., Ehlers, A., Clark, D.M., Tolin, D.F., & Orsillo, S.M. (1999). The posttraumatic cognitions inventory (PTCI): Development and validation. *Psychological Assessment 11(3)*, 303-314.

Folkman, S., & Lazarus, R. S. (1988). *Ways of coping questionnaire, permission set, manual, test booklets, scoring key.* Consulting Psychologists Press, Inc.

Fothergill, C.D., & Kuyken, W. (2002). *The quality of cognitive case formulation rating scale.* Unpublished manuscript.

Fried, E. I., van Borkulo, C. D., Cramer, A. O. J., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: A review of recent insights. *Social Psychiatry and Psychiatric Epidemiology*, *52*(1), 1–10. https://doi.org/10.1007/s00127-016-1319-z

Ghaderi, A. (2006). Does individualization matter? A randomized trial of standardized (focused) versus individualized (broad) cognitive behavior therapy for bulimia nervosa. *Behaviour Research and Therapy*, *44*(2), 273–288. https://doi.org/10.1016/j.brat.2005.02.004

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., … Charney, D. S. (1989). The Yale-Brown Obsessive Compulsive Scale: I. Development, use, and reliability. *Archives of General Psychiatry*, *46*, 1006–1011.

Gower, P. (2011). *Therapist competence, case conceptualisation and therapy outcome in cognitive behavioural therapy*. Unpublished manuscript.

Granger, C.W. (1969). Investigating causal relations by econometric models and cross-spectral models. *Econometrica: journal of the Econometric Society*, 424-438.

Graziano, A.M., & Raulin, A.L. (2014). *Research methods: A process of enquiry* (8th ed.). Pearson.

Greene, T., Gelkopf, M., Epskamp, S., & Fried, E. (2018). Dynamic networks of PTSD symptoms during conflict. *Psychological Medicine*, *48*(14), 2409–2417. https://doi.org/10.1017/S0033291718000351

Greene, T., Gelkopf, M., Fried, E. I., Robinaugh, D. J., & Lapid Pickman, L. (2020). Dynamic network analysis of negative emotions and DSM-5 posttraumatic stress disorder symptom clusters during conflict. *Journal of Traumatic Stress*, *33*(1), 72–83. https://doi.org/10.1002/jts.22433Haarhoff, B. (2008). *The map, the navigator, and the explorer: evaluating the content and quality of CBT case conceptualisations and the role of self-practice/self-reflection as a training intervention.* [Unpublished doctoral thesis], Massey University, Auckland, New Zealand.

Haarhoff, B. A., Flett, R. A., & Gibson, K. L. (2011). Evaluating the content and quality of Cognitive-Behavioral Therapy case conceptualisations. *New Zealand Journal of Psychology*, *40*(3), 104–114.

Haddock, G., McCarron, J., Tarrier, N., & Faragher, E. B. (1999). Scales to measure dimensions of hallucinations and delusions. The psychotic symptom rating scales (PSYRATS). *Psychological Medicine, 29(4)*, 879–889.

Hallam, R.S. (2013). *Individual Case Formulation*. Elsevier.

Harper, D. & Moss, D. (2003). A different kind of chemistry? Reformulating 'formulation'. *Clinical Psychology, 25,* 6-10.

Haynes, S. N., O'Brien, W. H., & Kaholokula, J. K. (2011). *Behavioral assessment and case formulation*. Wiley.

Health & Care Professions Council. (2015). *The standards of proficiency for practitioner psychologists.* https://www.hcpc-uk.org/standards/standards-of-proficiency/practitioner-psychologists

Hess, S. M. (2001). *The effects of a case formulation approach on process and outcome in the treatment of depression* (Doctoral Thesis, University of Texas at Austin, USA). ProQuest Information & Learning. http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2001-95010-179&site=ehost-live

Horowitz, L., Rosenberg, S., Baer, B., Ureno, G., & Villasenor, V. S. (1988). The Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885-892.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of counseling psychology, 36(2)*, 223.

Kahlon, S., Neal, A., & Patterson, T. G. (2014). Experiences of cognitive behavioural therapy formulation in clients with depression. *The Cognitive Behaviour Therapist*, *7*. https://doi.org/10.1017/S1754470X14000075

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist, 58*, 697–720.

Kendjelic, E. M., & Eells, T. D. (2007). Generic psychotherapy case formulation training improves formulation quality. *Psychotherapy: Theory, Research, Practice, Training*, *44*(1), 66–77. https://doi.org/10.1037/0033-3204.44.1.66

Kennedy, P., & Llewelyn, S. (2001). Does the future belong to the scientist practitioner? *The Psychologist, 14(2)*, 74-78.

King, L., McKenzie-McHarg, K., Horsch, A. (2017). Teaching a cognitive model to predict posttraumatic stress disorder following childbirth, *BMC Pregnancy and Childbirth, 17*(32), 1-12. https://doi.org/10.1186/s12884-016-1194-3

Kleim, B., Ehlers, A., & Glucksman, E. (2012). Investigating cognitive pathways to psychopathology: Predicting depression and posttraumatic stress disorder from early responses after assault. *Psychological Trauma: Theory, Research, Practice and Policy, 4*(5), 527-537. https://doi.org/10.1037/a0027006

Koerner, K. D. (1997). *The reliability and validity of a cognitive-behavioral case formulation method* (Doctoral Thesis, University of Washington, USA). ProQuest Information & Learning. http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1997-95024-148&site=ehost-live

Kramer, I., Simons, C. J. P., Wigman, J. T. W., Collip, D., Jacobs, N., Derom, C., Thiery, E., van Os, J., Myin-Germeys, I., & Wichers, M. (2014). Time-lagged moment-to-moment interplay between negative affect and paranoia: New insights in the affective pathway to psychosis. *Schizophrenia Bulletin*, *40*(2), 278–286. https://doi.org/10.1093/schbul/sbs194

Kramer, U., Vaudroz, C., Ruggeri, O., & Drapeau, M. (2013). Biased thinking assessed by external observers in borderline personality disorder. *Psychology and Psychotherapy: Theory, Research and Practice*, *86*(2), 183–196. https://doi.org/10.1111/j.2044-8341.2011.02056.x

Kroeze, R., Veen, D. C. van der, Servaas, M. N., Bastiaansen, J. A., Voshaar, R. C. O., Borsboom, D., Ruhe, H. G., Schoevers, R. A., & Riese, H. (2017). Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study. *Journal for Person-Oriented Research*, *3*(1), 1–10. https://doi.org/10.17505/jpor.2017.01

Kühberger, A. (1998). The Influence of Framing on Risky Decisions: A Meta-analysis. *Organizational Behavior and Human Decision Processes*, *75*(1), 23–55. https://doi.org/10.1006/obhd.1998.2781

Kuyken, W., Fothergill, C. D., Musa, M., & Chadwick, P. (2005). The reliability and quality of cognitive case formulation. *Behaviour Research and Therapy*, *43*(9), 1187–1201. https://doi.org/10.1016/j.brat.2004.08.007

Kwon, S.-M., & Oei, T. P. S. (1994). The roles of two levels of cognitions in the development, maintenance, and treatment of depression. *Clinical Psychology Review*, *14*(5), 331–358. https://doi.org/10.1016/0272-7358(94)90032-9

Lambert, M. J., Hansen, N. B., Umpress, V., Lunnen, K., Okiishi, J., & Burlingame, G. M. (1996). In M.D. Stevenson, *Administration and scoring manual for the OQ-45.2 (outcome questionnaire).* American Professional Credentialing Services LLC.

Lauritzen, S.L. (1996). *Graphical models (Vol. 17).* Clarendon Press.

Leahy, R.L. & Beck, A.T. (1996). Cognitive therapy of depression and mania. In R. Leahy (Ed.) *Cognitive therapy: Basic principles and applications*. (pp. 129-150). Jason Aronson.

Lipman, R. S., Covi, L., & Shapiro, A. K. (1979). The Hopkins symptom checklist (HSCL). *Journal of Affective Disorders, 1*, 9–24.

McConnaughy, E. A., DiClemente, C. C., Prochaska, J. O., & Velicer, W. F. (1989). Stages of change in psychotherapy: A follow-up report. *Psychotherapy: Theory, Research, Practice, Training, 26(4)*, 494.

McNally, R. J., Heeren, A., & Robinaugh, D. J. (2017). A Bayesian network analysis of posttraumatic stress disorder symptoms in adults reporting childhood sexual abuse. *European Journal of Psychotraumatology*, *8*(sup3), 1341276. https://doi.org/10.1080/20008198.2017.1341276

McNally, R. J., Robinaugh, D. J., Wu, G. W. Y., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*, *3*(6), 836–849. https://doi.org/10.1177/2167702614553230

Mumma, G. H. (2004). Validation of Idiosyncratic Cognitive Schema in Cognitive Case Formulations: An Intraindividual Idiographic Approach. *Psychological Assessment*, *16*(3), 211–230. https://doi.org/10.1037/1040-3590.16.3.211

Mumma, G.H. (2011). Validity issues in cognitive behavioral case formulation. *European Journal of Psychological Assessment, 27(1),* 29-49. https://doi.org/10.1027/1015-5759/a000054

Mumma, G. H., & Fluck, J. (2016). How valid is your case formulation? Empirically testing your cognitive behavioural case formulation for tailored treatment. *The Cognitive Behaviour Therapist*, *9*. https://doi.org/10.1017/S1754470X16000088

Mumma, G. H., Marshall, A. J., & Mauer, C. (2018). Person-specific validation and testing of functional relations in cognitive-behavioural case formulation: Guidelines and options. *Clinical Psychology & Psychotherapy*, *25*(5), 672–691. https://doi.org/10.1002/cpp.2298

Mumma, G. H., & Mooney, S. R. (2007). Comparing the validity of alternative cognitive case formulations: A latent variable, multivariate time series approach. *Cognitive Therapy and Research*, *31*(4), 451–481. https://doi.org/10.1007/s10608-006-9062-5

Mumma, G. H., & Smith, J. L. (2001). Cognitive-behavioral-interpersonal scenarios: Interformulator reliability and convergent validity. *Journal of Psychopathology and Behavioral Assessment*, *23*(4), 203–221. https://doi.org/10.1023/A:1012738802126

Muran, J. C. (1993). The self in cognitive-behavioral research: An interpersonal perspective. *The Behavior Therapist, 16*, 69-73.

Muran, J., Samstag, L., Segal, Z., & Winston, A. (1998). Interpersonal Scenarios: An Idiographic Measure of Self-Schemas. *Psychotherapy Research*, *8*(3), 321–333. https://doi.org/10.1080/10503309812331332417

Muran, J. C, Samstag, L.W., Ventur, E.D., Segal, Z.V., & Winston, A. (2001). A cognitive-interpersonal case study of a self. *Journal of Clinical Psychology, 57(3)*, 307-330. https://doi.org/10.1002/jclp.1014#

Muran, J. C., & Segal, Z. V. (1992). The development of an idiographic measure of self-schemas: An illustration of the construction and use of self-scenarios. *Psychotherapy: Theory, Research, Practice, Training*, *29*(4), 524–535. https://doi.org/10.1037/0033-3204.29.4.524

Muran, J. C., Segal, Z.V., & Samstag, L.W. (1994). Self-scenarios as a repeated measures outcome measurement of self-schemas in short-term cognitive therapy. *Behavior Therapy*, *25*(2), 255–274. https://doi.org/10.1016/S0005-7894(05)80287-4

Murray, J., Ehlers, A., & Mayou, R.A. (2002). Dissociation and post-traumatic stress disorder: Two prospective studies of road traffic accident survivors. *British Journal of Psychiatry, 180*, 363-368.

National Institute for Health and Care Excellence (2009) *Depression in adults: recognition and management* (NICE Clinical Guideline CG90) https://www.nice.org.uk/guidance/cg90

National Institute for Health and Care Excellence (2011) *Generalised anxiety disorder and panic disorder in adults: management* (NICE Clinical Guideline CG113) https://www.nice.org.uk/guidance/cg113

National Institute for Health and Care Excellence (2005) *Obsessive-compulsive disorder and body dysmorphic disorder: treatment* (NICE Clinical Guideline CG31) https://www.nice.org.uk/guidance/cg31

Nattrass, A., Kellett, S., Hardy, G. E., & Ricketts, T. (2015). The content, quality and impact of cognitive behavioural case formulation during treatment of obsessive compulsive disorder.

Behavioural and Cognitive Psychotherapy, *43*(5), 590–601. https://doi.org/10.1017/S135246581400006X

Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J., & Seller, R. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, *49*(1), 47–53. https://doi.org/10.1016/j.ijnurstu.2011.07.002

Pain, C. M., Chadwick, P., & Abba, N. (2008). Clients' experience of case formulation in cognitive behaviour therapy for psychosis. *British Journal of Clinical Psychology*, *47*(2), 127–138. https://doi.org/10.1348/014466507X235962

Persons, J.B. (1989). *Cognitive therapy in practice: A case formulation approach*. W.W. Norton & Company.

Persons, J. B., & Bertagnolli, A. (1999). Inter-rater reliability of cognitive–behavioral case formulations of depression: A replication. *Cognitive Therapy and Research*, *23*(3), 271–283. https://doi.org/10.1023/A:1018791531158

Persons, J. B., Bostrom, A., & Bertagnolli, A. (1999). Results of Randomized Controlled Trials of Cognitive Therapy for Depression Generalize to Private Practice. *Cognitive Therapy and Research*, *23*(5), 535–548. https://doi.org/10.1023/A:1018724505659

Persons, J. B., & Bums, D. D. (1985). Mechanism of action of cognitive therapy: Relative contribution of technical and interpersonal intervention. *Cognitive Therapy and Research, 9*, 539-551.

Persons, J.B., & Hong, J.J. (2016). Case formulation and the outcome of cognitive behaviour therapy. In N. Tarrier & J. Johnson (Eds.), *Case Formulation in Cognitive Behaviour Therapy*. (2nd edition, 58-109). Routledge.

Persons, J. B., Mooney, K. A., & Padesky, C. A. (1995). Interrater reliability of cognitive-behavioral case formulations. *Cognitive Therapy and Research*, *19*(1), 21–34. https://doi.org/10.1007/BF02229674

Persons, J. B., Roberts, N. A., Zalecki, C. A., & Brechwald, W. A. G. (2006). Naturalistic outcome of case formulation-driven cognitive-behavior therapy for anxious depressed outpatients. *Behaviour Research and Therapy*, *44*(7), 1041–1051. https://doi.org/10.1016/j.brat.2005.08.005

Persons, J.B., & Tompkins, M.A. (2007). Cognitive-behavioral case formulation. In T.D. Eells (Ed.), *Handbook of psychotherapy case formulation* (2nd ed. pp. 290-316). The Guilford Press.

Piccirillo, M. L., & Rodebaugh, T. L. (2019). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clinical Psychology Review*, *71*, 90–100. https://doi.org/10.1016/j.cpr.2019.01.002

Pluye P., & Hong, Q.N. (2014). Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. *Annual Review of Public Health, 35(1),* 29-45. https://doi-org.ezproxy01.rhul.ac.uk/10.1146/annurev-publhealth-032013-182440

Procidano, M. E., & Heller, K. (1983). Measures of perceived social support from friends and from family: Three validation studies. *American Journal Community Psychology, 11*, 1–24.

Rapee, R. M. (1991). Generalized anxiety disorder: A review of clinical features and theoretical concepts. *Clinical Psychology Review*, *11*(4), 419–440. https://doi.org/10.1016/0272-7358(91)90116-C

Redhead, S., Johnstone, L., & Nightingale, J. (2015). Clients' experiences of formulation in cognitive behaviour therapy. *Psychology and Psychotherapy: Theory, Research and Practice*, *88*(4), 453–467. https://doi.org/10.1111/papt.12054

Reeves, J.W., & Fisher, A.J. (2020). An examination of idiographic networks of posttraumatic stress disorder symptoms. *Journal of Traumatic Stress, 33*, 84-95. https://doi.org/10.1002/jts.22491

Regambal, M.J., & Alden, L.E. (2012). The contribution of threat probability estimates to reexperiencing symptoms: A prospective analog study. *Journal of Behavior Therapy and Experimental Psychiatry, 43(3)*, 947-951. https://doi.org/10.1016/j.jbtep.2012.02.003

Robinaugh, D., Haslbeck, J., Waldorp l., Kossakowski, J., Fried, E., Millner, A., McNally, R.J., van Nes, E.H., Scheffer, M., Kendler, K.S., & Borsboom, D. (2019). Advancing the network theory of mental disorders: A computational model of panic disorder. *Preprints.* Advance online publication. www.psyarxiv.com

Robinaugh, D.J., Hoekstra, R.H.A., Toer, E.R., & Borsboom, D. (2020). The network approach to psychopathology: a review of the literature 2008-2018 and an agenda for future research. Psychological Medicine, 50, 353-366. https://doi.org/10.1017.S0033291719003404

Rosenberg, M. (1979). *Conceiving the self.* Basic Books, Inc.

Rosenberg, S., Silberschatz, G., Curtis, J., Sampson, H., & Weiss, J. (1986). The plan diagnosis method: A new approach to establishing reliability for psychodynamic formulations. American Journal of Psychiatry, 143(11), 1454-1456.

Roth, D. A., & Heimberg, R. G. (2001). COGNITIVE-BEHAVIORAL MODELS OF SOCIAL ANXIETY DISORDER. *Psychiatric Clinics of North America*, *24*(4), 753–771. https://doi.org/10.1016/S0193-953X(05)70261-6

Santangelo, P.S., Ebner-Priemer, U.W., and Trull, T.J. (2013). Experience sampling methods in clinical psychology. *The Oxford Handbook of Research Strategies for Clinical Psychology*, 188–210.

Schumacher, L., Burger, J., Zoellner, F., Zindler, A., Epskamp, S., & Barthel, D. (2021). Using clinical expertise and empirical data in constructing networks of trauma symptoms in refugee youth. *European Journal of Psychotraumatology*, *12*(1), 1920200. https://doi.org/10.1080/20008198.2021.1920200

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, pp. 420-428.

Spengler, P. M., & Strohmer, D. C. (1994). Clinical judgmental biases: The moderating roles of counselor cognitive complexity and counselor client preferences. *Journal of Counseling Psychology*, *41*(1), 8–17. https://doi.org/10.1037/0022-0167.41.1.8

Spiller, T. R., Schick, M., Schnyder, U., Bryant, R. A., Nickerson, A., & Morina, N. (2017). Symptoms of posttraumatic stress disorder in a clinical sample of refugees: A network analysis. *European Journal of Psychotraumatology*, *8*(sup3), 1318032. https://doi.org/10.1080/20008198.2017.1318032

Stiles, W. B. (1980). Measurement of the impact of psychotherapy sessions. *Journal of Consulting and Clinical Psychology, 48*, 176-185.

Stone, A.A., & Shiffman, S. (1994) Ecological momentary assessment in behavioral medicine. *Annals of behavioral medicine, 16*, 199-202. https://doi.org/10.1093/abm/16.3.199

Teasdale, J. D. (1983). Negative thinking in depression: Cause, effect, or reciprocal relationship? *Advances in Behaviour Research and Therapy*, *5*(1), 3–25. https://doi.org/10.1016/0146-6402(83)90013-9

The Information Standard. (2013). *Finding the evidence*. tis-guide-finding-the-evidence-07nov.pdf

Touyz, S.W. (1995). Clinical psychology in disarray: The challenge ahead. *Australian Psychologist, 30*, 191-195.

Van der Krieke, L., Emerencia, A.C., Bos, E.H., Rosmalen, J.G., Riese, H., Aiello, M., Sytema, S., & de Jonge, P. (2015). Ecological momentary assessments and automated time series analysis to promote tailored health care: A proof-of-principle study. *JMIR research proposals, 4(3)*, e100. https://doi.org/10.2196/resprot.4000

Vrijen, C., Hartman, C.A., van Roekel, E., de Jonge, P., & Oldehinkel, A.J. (2018). Spread the joy: How high and low bias for happy facial emotions translate into different daily life affect dynamics. *Complexity, Article ID 2674523*, 1-15. https://doi.org/10.1155/2018/2674523

Weathers, F.W., Litz, B.T., Keane, T.M., Palmieri, P.A., Marx, B.P., & Schnurr, P.P. (2013). The PTSD Checklist for DSM-5 (PCL-5). Scale available from the National Center for PTSD at www.ptsd.va.gov

Wells, S.Y., Morland, L.A., Torres, E.M., Kloezeman, K., Mackintosh, M.-A., & Aarons, G.A. (2019). The development of a brief version of the posttraumatic cognitions inventory (PTCI-9). *Assessment, 26(9)*, 193-208. https://doi.org/10.1177/1073191116685401

Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, *44*(7), 1349–1360. https://doi.org/10.1017/S0033291713001979

Wichers, M., Wigman, J.T.W., Bringmann, L.F., & de Jonge, P. (2017). Mental disorders as networks: some cautionary reflections on a promising approach. *Social Psychiatry and Psychiatric Epidemiology, 52*(2), 143-145.

Wichers, M., Wigman, J.T.W., & Myin-Germeys, I. (2015). Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Review, 7(4)*, 362-367. https://doi.org/10.1177/1754073915590623

Wild, B., Eichler, M., Friederich, H.C., Hartmann, M., Zipfel, S., & Herzog, W. (2010). A graphical vector autoregressive modelling approach to the analysis of electronic diary data. *BMC medical research methodology, 10(1)*, 1-13. https://doi.org/10.1186/1471-2288-10-28

Wild, J., Warnock-Parkes, E., Grey, N., Stott, R., Wiedemann, M., Canvin, L., Rankin, H., Shepherd, E., Forkert, A., Clark, D.M., & Ehlers, A. (2016). Internet-delivered cognitive therapy for PTSD: a development pilot series. *European Journal of Psychotraumatology, 7*: 31019. https://doi.org/10.3402.ejpt.v7.31019

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67*, 361–370.

Appendices

Appendix A: Ethical Approval for the Study

NHS
Health Research Authority

South Central - Berkshire B Research Ethics Committee
The Old Chapel
Royal Standard Place
Nottingham
NG1 6FS

Tel: 02071048357

07 October 2020

Gary Brown
Research Director
Doctorate in Clinical Psychology
Royal Holloway University of London
Egham, Surrey
TW20 0EX

Dear Dr Brown

Study title:	**Predicting patterns of exacerbation and improvement in psychological therapies**
REC reference:	**17/SC/0204**
Amendment number:	**Amendment 3.0**
Amendment date:	**11 September 2020**
IRAS project ID:	**225649**

Thank you for submitting the above amendment, which was received on 28 September 2020. It is noted that this is a modification of an amendment previously rejected by the Committee (our letter of 02 September 2020 refers).

The modified amendment was reviewed by the Sub-Committee in correspondence. A list of the members who took part in the review is attached.

Ethical opinion

As part of the review the sub-committee made a number of suggestions regarding the protocol and participant information sheet and thank you for providing revised documents.

I am pleased to confirm that the Committee has given a favourable ethical opinion of the modified amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved are:

Document	*Version*	*Date*
Covering letter on headed paper [modified amendment]	n/a	18 September 2020
Non-validated questionnaire [PCTI questionnaire]	n/a	18 September 2020
Non-validated questionnaire [RIQ-Questionnaire]	n/a	18 September 2020

Non-validated questionnaire [PCL-5]	n/a	18 September 2020
Notice of Modified Amendment	3.0	11 September 2020
Other [5 Research Proposal - Clean version - SCED + App]	1.7	06 October 2020
Participant information sheet (PIS) [clean]	v1.5	18 September 2020
Participant information sheet (PIS) [for app subsample clean]	1.6	06 October 2020
Summary CV for student	n/a	21 July 2020
Summary CV for student	n/a	

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at: https://www.hra.nhs.uk/planning-and-improving-research/learning/

IRAS Project ID - 225649:	**Please quote this number on all correspondence**

Yours sincerely

Darren Fletcher
On behalf of

Dr John Sheridan
Chair

E-mail: berkshireb.rec@hra.nhs.uk

Enclosures: *List of names and professions of members who took part in the review*

Copy to: *Elizabeth Blomfield, Liza Turner*

Health Research Authority

South Central - Berkshire B Research Ethics Committee

Attendance at Sub-Committee of the REC meeting on 05 October 2020

Committee Members:

Name	*Profession*	*Present*	*Notes*
Dr Alan Clark	Pharmacist	Yes	
Dr John Sheridan	Consultant Toxicologist and Chemist	Yes	

Also in attendance:

Name	*Position (or reason for attending)*
Mia Cooper	Approvals Administrator

Appendix B: Participant Information Sheet

Mechanisms of change in psychological therapy

Participant Information Sheet

You are being asked to allow information from questionnaires you complete during therapy to be used in a research study. This study is being conducted as part of a Clinical Psychology Doctorate Thesis. This study is being run at various London psychotherapy services, including the one at which you are being seen.

Why have I been invited to take part?

You have been invited to take part as you have attended an assessment appointment and are currently either on the waiting-list or have started to receive psychological therapy from a participating service.

It is known that psychological therapy helps to improve symptoms for a number of different mental health difficulties. However, we want to look in more detail at factors that might predict outcomes in psychological therapy by looking in more detail as what changes over time as people get better.

What will I have to do?

Whilst you are on the waiting-list and during your therapy, your therapist will ask you to complete some questionnaires. Everyone who has an assessment or receives treatment from a psychotherapy service is asked to complete questionnaires to help understand how they are feeling and to look at changes during therapy.

If you are eligible for the study and happy to participate, you will be asked to complete three questionnaires before joining the study. These are brief and should take no more than 15 minutes.

Based on the responses to your assessment, the study researcher will draw up a personalized questionnaire consisting of fewer than ten questions. During your time in the study, you will be filling out these personalized questionnaires daily through a browser on your phone that will connect to a website called “Qualtrics” that will record your answers. The researcher will help you become familiar with the process and practice completing the forms online.

When you start the study, you will be asked to start filling out your personal questionnaires five times a day. You will be reminded to do this by an SMS text message which links to the questionnaire and will have two hours during which to respond. You will complete these questionnaires for two weeks of therapy.

After these two weeks, you will be asked to complete three final general questionnaires and the results will be shared with your clinician to help plan your ongoing treatment.

To compensate you for your time, you will receive payment for completing the personal questionnaires you receive on your phone. You will be paid £1 per questionnaire completed plus an additional £1 daily bonus for completing all surveys received that day. This payment will be calculated and given to you after you complete your final general questionnaires.

Do I have to take part?

No, it is completely up to you. Your decision will not affect the healthcare you receive in any way.

Are there any benefits for me?

There are unlikely to be any direct benefits to you from taking part in the study. You are currently receiving treatment from an NHS service, and there won't be any changes to the treatment you receive through taking part in this study. We hope that this study will help us to understand more about psychological therapy and how it works, and be of benefit in the future.

Are there any risks for me?

There are no risks involved in taking part in this study as we are using information collected as part of routine practice and the same information will be collected between sessions. If you feel uncomfortable or concerned about any of the questionnaires, your therapist will be able to talk about this with you and will only continue if you are happy to do so.

How will be use information about you?

We will need to use information from you and from your medical records for this research project.

This information will include

- your email address and mobile phone number.

People will use this information to do the research or to check your records to make sure that the research is being done properly.

People who do not need to know who you are will not be able to see your name or contact details. Your data will have a code number instead.

We will keep all information about you safe and secure.

Once we have finished the study, we will keep some of the data so we can check the results. We will write our reports in a way that no-one can work out that you took part in the study.

What are your choices about how your information is used?

- You can stop being part of the study at any time, without giving a reason, but we will keep information about you that we already have.

Where can you find out more about how your information is used?

You can find out more about how we use your information

- at www.hra.nhs.uk/information-about-patients/
- our leaflet available from **www.hra.nhs.uk/patientdataandresearch**
- by asking one of the research team by sending an email to dataprotection@royalholloway.ac.uk

All data entered on the Qualtrics website is stored anonymously using your code number. The email address you provide will be stored in a file separate from your responses. This file will only be kept until you have finished the final set of questions. Once you no longer need to be contacted to remind you to complete your forms, your email address will be deleted from the Qualtrics website.

Who has approved the study?

All research in the NHS is reviewed by an independent group of people, called a Research Ethics Committee, which is there to protect your safety, wellbeing, rights and dignity. This project has been reviewed and was given a favourable review by the South Central – Berkshire B Research Ethics Committee on 18/02/21.

What happens next?

If you are willing for your data to be used for this research study, please let your therapist know.

What if something goes wrong?

It is unlikely that you will encounter any difficulties in the course of your participation. However, should you wish to contact an independent office to discuss any difficulties or submit a complaint, please contact [PALS information]

Further information and contact details

If you would like any further information, please contact Elizabeth Blomfield (e.blomfield@nhs.net and Elizabeth.Blomfield.2017@live.rhul.ac.uk).

Thank you for taking the time to read this information and for your interest in our research.

Appendix C: Consent Form

IRAS ID: 225649

Centre Number:

Study Number: 1

Participant Identification Number for this trial:

CONSENT FORM

Title of Project: Predicting patterns of exacerbation and improvement in psychological therapies

Name of Researcher: Elizabeth Blomfield

Please initial box

1. I confirm that I have read the information sheet dated..... for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

3. I understood that I will receive payment of £6 per day on which I completed in full five personalized questionnaires delivered through the Qualtrics website on my phone (£1 per survey completed plus £1 per day when all surveys completed). I agree that I will receive this payment after completing the three final general questionnaires. ☐

4. I understand that relevant sections of my medical notes and data collected during the study, may be looked at by individuals from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐

5. I agree to my General Practitioner being informed of my participation in the study. ☐

6. I agree to take part in the above study. ☐

Name of Participant	Date	Signature
Name of Person taking consent	Date	Signature

Appendix D: PCL-5 Questionnaire (Weathers et al., 2013)

PCL-5

Instructions: Below is a list of problems that people sometimes have in response to a very stressful experience. Please read each problem carefully and then circle one of the numbers to the right to indicate how much you have been bothered by that problem in the past month.

In the past month, how much were you bothered by:	Not at all	A little bit	Moderately	Quite a bit	Extremely
1. Repeated, disturbing, and unwanted memories of the stressful experience?	0	1	2	3	4
2. Repeated, disturbing dreams of the stressful experience?	0	1	2	3	4
3. Suddenly feeling or acting as if the stressful experience were actually happening again (as if you were actually back there reliving it)?	0	1	2	3	4
4. Feeling very upset when something reminded you of the stressful experience?	0	1	2	3	4
5. Having strong physical reactions when something reminded you of the stressful experience (for example, heart pounding, trouble breathing, sweating)?	0	1	2	3	4
6. Avoiding memories, thoughts, or feelings related to the stressful experience?	0	1	2	3	4
7. Avoiding external reminders of the stressful experience (for example, people, places, conversations, activities, objects, or situations)?	0	1	2	3	4
8. Trouble remembering important parts of the stressful experience?	0	1	2	3	4
9. Having strong negative beliefs about yourself, other people, or the world (for example, having thoughts such as: I am bad, there is something seriously wrong with me, no one can be trusted, the world is completely dangerous)?	0	1	2	3	4
10. Blaming yourself or someone else for the stressful experience or what happened after it?	0	1	2	3	4
11. Having strong negative feelings such as fear, horror, anger, guilt, or shame?	0	1	2	3	4
12. Loss of interest in activities that you used to enjoy?	0	1	2	3	4
13. Feeling distant or cut off from other people?	0	1	2	3	4
14. Trouble experiencing positive feelings (for example, being unable to feel happiness or have loving feelings for people close to you)?	0	1	2	3	4
15. Irritable behavior, angry outbursts, or acting aggressively?	0	1	2	3	4
16. Taking too many risks or doing things that could cause you harm?	0	1	2	3	4
17. Being "superalert" or watchful or on guard?	0	1	2	3	4
18. Feeling jumpy or easily startled?	0	1	2	3	4
19. Having difficulty concentrating?	0	1	2	3	4
20. Trouble falling or staying asleep?	0	1	2	3	4

Appendix E: Posttraumatic Cognitions Inventory Questionnaire (PTCI; Foa et al., 1999)

posttraumatic cognitions inventory (pcti)

your name: *today's date:*

We are interested in the kind of thoughts which you may have had after a traumatic experience. Below are a number of statements that may or may not be representative of your thinking. Please read each statement carefully and tell us how much you AGREE or DISAGREE with each by putting the appropriate number between 1 & 7 in the box to the right of the statement. People react to traumatic events in many different ways. There are no right or wrong answers to these statements.

1	2	3	4	5	6	7
totally disagree	*disagree very much*	*disagree slightly*	*neutral*	*agree slightly*	*agree very much*	*totally agree*

1.	the event happened because of the way I acted	
2.	I can't trust that I will do the right thing	
3.	I am a weak person	
4.	I will not be able to control my anger and will do something terrible	
5.	I can't deal with even the slightest upset	
6.	I used to be a happy person but now I am always miserable.	
7.	people can't be trusted	
8.	I have to be on guard all the time	
9.	I feel dead inside	
10.	you can never know who will harm you	
11.	I have to be especially careful because you never know what can happen next	
12.	I am inadequate	
13.	if I think about the event, I will not be able to handle it	
14.	the event happened to me because of the sort of person I am	
15.	my reactions since the event mean that I am going crazy	
16.	I will never be able to feel normal emotions again	
17.	the world is a dangerous place	
18.	somebody else would have stopped the event from happening	
19.	I have permanently changed for the worse	
20.	I feel like an object, not like a person	
21.	somebody else would not have gotten into this situation	
22.	I can't rely on other people	
23.	I feel isolated and set apart from others	
24.	I have no future	
25.	I can't stop bad things from happening to me	
26.	people are not what they seem	
27.	my life has been destroyed by the trauma	
28.	there is something wrong with me as a person	
29.	my reactions since the event show that I am a lousy coper	
30.	there is something about me that made the event happen	
31.	I feel like I don't know myself anymore	
32.	I can't rely on myself	
33.	nothing good can happen to me anymore	

Appendix F: Response to Intrusions Questionnaire (Clohessy & Ehlers, 1999; Murray, Ehlers, & Mayou, 2002)

RIQ

What do you do when memories of the traumatic event pop into your mind? Please circle the answer that applied best to you DURING THE PAST MONTH.

1.	I try to push them out of my mind.	Never	Sometimes	Often	Always
2.	I try to erase the memory of the event.	Never	Sometimes	Often	Always
3.	I try hard to control my emotions.	Never	Sometimes	Often	Always
4.	I distract myself with something else.	Never	Sometimes	Often	Always
5.	I think of something else.	Never	Sometimes	Often	Always
6.	I work hard at keeping busy with other things.	Never	Sometimes	Often	Always
7.	I think about how life would have been different if the event had not occurred.	Never	Sometimes	Often	Always
8.	I dwell on how the event could have been prevented.	Never	Sometimes	Often	Always
9.	I think about why the event happened to me.	Never	Sometimes	Often	Always
10.	I dwell on how I used to be before the event.	Never	Sometimes	Often	Always
11.	I dwell on what other people have done to me.	Never	Sometimes	Often	Always
12.	I dwell on what I should have done differently.	Never	Sometimes	Often	Always
13.	I go over what happened again and again.	Never	Sometimes	Often	Always
14.	I worry that something similar will happen to me or my family.	Never	Sometimes	Often	Always
15.	I detach myself from the memories.	Never	Sometimes	Often	Always
16.	I drift off into a world of my own.	Never	Sometimes	Often	Always
17.	I numb my feelings.	Never	Sometimes	Often	Always
18.	I drink alcohol, take medication or use drugs.	Never	Sometimes	Often	Always
19.	I put on loud music or TV.	Never	Sometimes	Often	Always

Appendix G: Example of Personalised Participant Survey

Welcome! The time of this survey is….

1. Has something triggered your PTSD symptoms since you last responded?
   [Yes/No]

   1a. What was it that triggered your symptoms?
   [Open text]

2. Since you last responded, how much have you been bothered by… disturbing and unwanted memories of your stressful experience?
   [Not at all / A little bit / Moderately / Quite a bit / Extremely

3. Since you last responded, how much were you bothered by… suddenly feeling or acting as if your stressful experience was actually happening again (as if you were actually back there reliving it)?
   [Not at all / A little bit / Moderately / Quite a bit / Extremely]

4. Since you last responded, how much were you bothered by… feeling very upset or having strong physical reactions when something reminded you of your stressful experience?
   [Not at all / A little bit / Moderately / Quite a bit / Extremely]

5. Since you last responded, how much have you been bothered by…. being 'superalert' or watchful or on guard?
   [Not at all / A little bit / Moderately / Quite a bit / Extremely]

6. Since you last responded, how much have you been bothered by… feeling jumpy or easily startled?
   [Not at all / A little bit / Moderately / Quite a bit / Extremely]

7. How much do you agree or disagree with the following statements? (Top 3 PCTI scores)

   I am a weak person.
   I feel isolated and set apart from others.
   The world is a dangerous place.
   [Strongly disagree / Disagree / Neither agree nor disagree / Agree / Strongly agree]

8. Since you last responded, when memories of your stressful event popped into your mind, to what extent did you… (top 3 RiQ scores)

   I distract myself with something else.
   I dwell on what I should have done differently.
   I drift off into a world of my own.
   [Never /Sometimes/Often/Always]

9. Right now….(ideographic contextual moderator)
   …I am feeling tired.
   [Yes / No]

10. Right now… (idiographic contextual moderator)

…I am alone.
[Yes / No]

Appendix H: Example of Symptom Relationships Table

This box includes all the symptoms from your personalised survey listed horizontally and vertically. Please complete the table, below, by placing an X in any unshaded box where you feel any two symptoms tend to connect to each other in some way e.g. they usually occur together, or one often leads very quickly to another. For example, if you notice that feeling as if the stressful event is actually happening often leads you to think people can't be trusted, you would place a cross in the white box which connects those two symptoms – column 3, line 7c.

	1 Trigger	2 Disturbing and unwanted memories	3 Reliving the stressful experience (day or	4 Feeling very upset or having strong physical reactions	5 Being superalert watchful or on guard	6 Feeling jumpy or easily startled	7a Thought: There is something about me that made the event happen.	7b Thought: I will never be able to feel normal emotions again.	7c Thought: People can' t be trusted.	8a Response: I work hard at keeping busy with other things	8b Response: I worry that something similar will happen to me or my family.	8c Response: I detach myself from the	9 Right now… I am in an academic lecture / seminar.	10 Right now… I have slept less than 3 hours last night.	11 Right now… my sister is out socialising.
1. Has something triggered your PTSD symptoms since you last responded?															
2. Since you last responded, how much have you been bothered by **disturbing & unwanted memories of your stressful experience?**															
3. Since you last responded, how much were you bothered by…**suddenly feeling or acting as if your stressful experience was actually happening again (as if you were actually back there reliving it)?**															

4. Since you last responded, how much were you bothered by…**feeling very upset or having strong physical reactions when something reminded you of your stressful experience?**															
5. Since you last responded, how much have you been bothered by….**being 'superalert' or watchful or on guard?**															
6. Since you last responded, how much have you been bothered by…**feeling jumpy or easily startled?**															
7a. How much do you agree or disagree with the following statements?: **There is something about me that made the event happen.**															
7b. How much do you agree or disagree with the following statements? **I will never be able to feel normal emotions again.**															
7c. How much do you agree or disagree with the following statements? **People can't be trusted.**															
8a. Since you last responded, when memories of your stressful event popped into your mind, to what extent did you… **work hard at keeping busy with other things?**															
8b. Since you last responded, when memories of your stressful event popped into your mind, to what extent did you… **Worry that something similar will happen to you or your family?**															
8c. Since you last responded, when memories of your stressful event popped into your mind, to what extent did you… **detach yourself from the memories?**															
9. Right now… **I am in an academic lecture / seminar.**															

10. Right now… **I have slept less than 3 hours last night**.																
11. Right now… **my sister is out socialising.**																