A replicated, single case, feasibility study of group cognitive behavioural therapy+ for provoked vulvodynia.

Clover Giles

Örebro University

Supervisor: Johan Carstens-Söderstand

Master’s thesis

Professional Psychology Program

Autumn 2018
To Pomona and Matz

Pomona,
Thank you for giving perspective and endless love!
You’ve shown me that there are much more important and challenging things in life than a master’s thesis!

Matz,
Thank you for being there, all the way.
Without you there wouldn’t have been any psychologist program. I hope I’ve done you proud!

(Bellybean, thanks for not making me too nauseous!)
Abstract

Provoked vulvodynia is thought to be the most common form of vulvovaginal pain, affecting up to 8% of all cis-women. However, there are currently few treatment options. The primary aim of this thesis was to assess the feasibility of a group cognitive behavioural therapy (CBT) treatment, with couples’ sessions, for provoked vulvodynia. The secondary aim was to assess patterns of change and fit with the fear-avoidance model. The study was a single case experimental design and the sample consisted of five Scandinavian couples. Outcome variables were pain, sexual function and partner response. Secondary measures were compliance, completion, perceived credibility, depression and anxiety. Process variables were catastrophizing, fear and avoidance. Analysis was performed visually and statistically; using the percentage of data points exceeding the median (PEM) and Fisher’s exact test.

Weekly measures for pain showed ambiguous results and effect sizes ranged from small to moderate. Post treatment measures showed that pain was meaningfully reduced for 4 of 5 women. One woman reported deterioration. Improvements in pain were retained at 3 month follow-up. Weekly measures of function were also ambiguous, however slight improvements were seen in post treatment measures. Deterioration was observed in partner response. Compliance, completion and perceived credibility were good to excellent, but no clear effects were observed on depression and anxiety. Weekly measures of process variables failed to support the pattern of the fear-avoidance model. The implications of these results are, that although showing signs of promise, the treatment protocol needs refinement. Furthermore, to aid development of more effective pain treatments, future research is recommended to continue critically evaluating putative process measures such as catastrophizing, fear, and avoidance, and the pattern of the fear-avoidance model.

Key words: Provoked vulvodynia, pain, cognitive behavioural therapy, CBT, fear-avoidance model


Nyckelord: Provocerad vulvodyni, smärta, kognitiv beteende terapi, KBT, rädsla-undvikandemodellen
A replicated, single case, feasibility study of group cognitive behavioural therapy + for provoked vulvodynia.

Ten to 30% of all cis-women under 30 regularly suffer vulvovaginal pain during sexual intercourse (e.g., Thomtén, Lundahl, Stigenberg, & Linton, 2014; van Lankveld et al., 2010). A little over half of these report seeking treatment (e.g., Harlow et al., 2014). Current treatment methods range from surgery to topical ointments and psychological interventions, but of those who do seek treatment, many go 7 years or more before being successfully treated (Pacik, 2014). It is therefore evident that there are deficits in knowledge and effective treatments in this area. These deficits could reflect an historically neglected research subject, compounded with gender biases in the treatment of pain (e.g., Hoffmann & Tarzian, 2001). However, systematic reviews reveal that current psychological treatments for pain are only moderately successful at reducing symptoms and restoring function (e.g., Allen & Williams, 2001; Kamper et al., 2015). Thus, there is a need for development in this area of healthcare.

However, over the past decades, pain treatment research has followed a relatively narrow theoretical track. Although several models of chronic pain have been developed (see Linton, 2013, for an overview) much of the current research focuses on the fear avoidance model (FAM), including research on vulvovaginal pain (e.g., Benoit-Piau et al., 2018; Curtin & Norris, 2017). The FAM parsimoniously explains the one-way, circular development of chronic pain and proposes only one return to function pathway through reduced fear and then confrontation (see Figure 1.). Accordingly, one aspect of early interventions for pain is fear reduction. However, exposure is currently treatment of choice for many presentations of chronic pain (e.g., Linton, 2013, p. 301). Exposure is believed to most effectively reduce fear by first challenging avoidance, i.e., through confrontation (Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014). This appears to challenge the FAM’s unidirectionality. Additionally, the FAM implies that components early in the cycle have catalysing effects.
However, an explanation of components’ risk weight in the development of chronic pain is lacking. As such, many current psychological treatments, including those for vulvovaginal pain, may be based on speculative premises and thus be suboptimal.

Vulvovaginal pain entails suffering for those affected. Reduced wellbeing, and quality of life (Desrochers, Bergeron, Khalifé, Dupuis, & Jodoin, 2009), unsatisfactory relationships (Smith & Pukall, 2011), and comorbid mood and anxiety disorders (Desrochers et al., 2009) are common complaints. The symptoms of vulvovaginal pain may restrict women’s choice of clothing, transport, and recreation (Bachmann et al., 2006; Thomtén et al., 2014). Women with chronic vulvovaginal pain are also less likely to conceive, while those who do are less likely to labour or birth vaginally (Veasley & Witkin, 2015). Vulvovaginal pain is usually described as an excruciating burning sensation, often occurring when pressure is applied to the vulva. Pain may also be cutting or aching or unprovoked (Bergeron, Binik, Khalife, Pagidas, & Glazer, 2001). Pain is most often felt in or around the vaginal opening but can be felt in the whole vulva. “Deep pain” may also be felt within the vagina. Muscle spasms or tension in the pelvic floor may also be present, making penetration impossible for some (e.g., Pacik, 2014). Vulvovaginal pain thereby entails not only suffering, but also extensive functional disability.

**Provoked Vulvodynia**

All forms of vulvovaginal pain without clear medical explanations are currently gathered in the “Genito-Pelvic Pain/Penetration Disorder” (GPPD) diagnosis in the fifth edition of the American Psychiatric Association’s Diagnostic and Statistical Manual (DSM-V) (American Psychiatric Association, 2013). However, in previous editions of the manual, specific presentations of pain were divided into unique diagnoses. Provoked Vulvodynia (PVD) is one such subtype of vulvovaginal pain, for which treatments are currently being researched. According to the 2015 terminology of the International Society for the Study of
Vulvovaginal Disease (ISSVD), “vulvodynia is vulvovaginal pain without apparent biomedical cause, lasting at least three months. Vulvodynia may be specified as; localized, generalized or mixed; provoked, spontaneous or mixed; with primary or secondary onset; constant, intermittent, persistent, immediate or delayed.” (Bornstein, et al., 2016).

PVD is, as such, a form of vulvodynia. More specifically, PVD is vulvodynia without clearly identifiable cause, occurring when pressure is applied to the vulva, and recurring over at least three months. While roughly a third of women may suffer from some form of vulvovaginal pain (Thomtén et al., 2014), the prevalence of PVD in two urban U.S. populations was found to be around 8% (Harlow et al., 2014). PVD is thought to be the most common subtype of medically unexplained vulvovaginal pain (Reed et al., 2011). Thus, specific research attention is motivated.

**Treatment of Vulvovaginal Pain**

In a review and meta-analysis, medical and psychological treatments for vulvovaginal pain were found to have similar effects on symptom reduction (Flanagan, Herron, O'Driscoll, & Williams, 2015). However, psychological treatments proved to be equally effective regardless of symptom etiology. Examples of psychological treatment modalities recently trialled are: Mindfulness based group therapies (e.g., Brotto, Basson, Smith, Driscoll & Sadownik, 2014), couples therapy (e.g., Corsini-Munt, Bergeron, Rosen, Mayrand, & Delisle, 2014), exposure therapy (e.g., ter Kuile, Melles, Groot, Tuijman-Raasveld, van Lankveld, 2013), and group cognitive behavioural therapy (CBT) (ter Kuile & Weijenborg, 2006).

Mindfulness based CBT for PVD was tested (N = 85, 4 sessions) in an RCT by Brotto et al. (2014). The treatment group reported significant improvements in pain self-efficacy, catastrophizing, hypervigilance, sex related distress and pain. Effect sizes were not reported. Pain at FU correlated only with pain and number of comorbid pain complaints at inclusion, and changes in pain self-efficacy. In a small-scale trial of couples’ therapy for PVD (N=9, 12
sessions) Corsini-Munt et al. (2014) found significant improvements in women’s pain and function. Both partners’ sexual satisfaction and pain catastrophizing were also improved. Effect sizes for pain reductions varied from small to huge depending on the measures used. Moderate effect sizes were seen in sexual function and very large for sexual satisfaction and pain catastrophizing. CBT group therapy (N = 117, 10 x 2 hour sessions) for lifelong vaginismus was compared to a waiting list control and bibliotherapy by ter Kuile and Weijenborg (2006). Pain was not an outcome variable, but at post treatment sexual function was seen to have improved. Effect sizes were small. Another RCT of exposure therapy (N = 70, 3 x 2 hour sessions) for vaginismus was conducted by ter Kuile et al. (2013). Improvements were found in pain, intercourse frequency, vaginismus symptoms and catastrophizing. All effect sizes in comparison to the waitlist were very large. The treatment trialled in this study is based on the protocol by ter Kuile and Weijenborg (2006) which included psychoeducation, relaxation, sensate focus, exposure and cognitive restructuring.

This shows that several promising treatments for vulvovaginal pain are being developed. However, there is currently a lack of studies in Scandinavian contexts, and few repetitions to further support promising results. Moreover, earlier research has not investigated why, or the specific ways, in which interventions have effect. Thus, researchers and clinicians must continue to rely exclusively on theory to further improve results.

There are research designs which may further the understanding of the treatment process. For example, Single Case Experimental Designs (SCED) use repeated measures to illuminate change over time (e.g., Schemer et al., 2018). This allows assessment of the potency of interventions, and whether interventions work as expected. Replications of data patterns across multiple participants is seen to increase the external validity of results (Wright et al., 2015). Furthermore, non-replications may increase understanding of why seemingly logical interventions do not attain large effect sizes (e.g., Khandker et al. 2011). Thus,
SCEDs may provide indications for refinement of treatment protocols. However, published treatment SCEDs for vulvovaginal pain are currently lacking, and thus reliance on existing theory when developing treatments is paramount.

**The Biopsychosocial Perspective**

An empirical knowledge base is needed to construct scientific theories. Historically, empirical research on vulvovaginal pain was dichotomous; biological or psychosocial (Desrochers, Bergeron, Landry, & Jodoin, 2008). However, this dichotomy has recently been discarded in favour of a biopsychosocial approach (Thomté & Linton, 2013). As no single explanation exists, known risk factors for the development of vulvovaginal pain can be categorized as biological, psychological and social. These factors are assumed to interact, contributing to the development and maintenance of pain. Thus, it is upon this biopsychosocial evidence base that treatments can be founded.

Some of the known risk factors for vulvovaginal pain are: early onset of menarche (<11 years [Harlow, Wise, & Stewart, 2001]), repeated infections (Arnold, Bachman, Rosen, Kelly & Rhoads, 2006) and comorbidity with other pain disorders (Bergeron et al., 2015). Hormonal changes related to oral contraceptives (Goldstein, Kim, Burrows, & Goldstein, 2015), and neurological pain sensitization (Hampson et al., 2013) have also been supported. Common psychological correlates are anxiety, pain catastrophizing, fear-avoidance, hypervigilance, low pain self-efficacy, and dysfunctional pain beliefs (Bornstein et al., 2016). Compared to controls, women with pre-existing mood or anxiety disorders are at greater risk of developing vulvovaginal pain. Similarly, women with vulvovaginal pain are at greater risk of developing post facto mood or anxiety disorders (Khandker et al., 2011). Pain on first tampon use (e.g., Harlow et al., 2001), childhood trauma accompanied by fear of future abuse (Landry & Bergeron, 2011), and fear of partner loss (e.g., Enlund Tuuvas & Lennartsson, 2018) are also known correlates. Additionally, social risk factors can be found on micro,
meso and macro levels. Solicitous partner responses have been identified as a maintenance factor (Rosen, Bergeron, Lambert, & Steben, 2013), and associated with greater pain intensity (Flink, Engman, Thomtén, & Linton, 2017). Vulvovaginal pain has also been found to more prevalent among ethnic minorities (Harlow et al, 2014), and young Swedish women have reported enduring pain as their sexual ideals included penetration (Elmerstig, Wijma, & Berterö, 2008). There is, thus, support for the biopsychosocial perspective.

The risk factors described are not an exhaustive list, but they cumulatively illustrate the complex etiologies vulvovaginal pain may have; and, moreover, why treatment choices are not obvious. A theoretically structured approach is needed to establish relationships between risk factors. Furthermore, indications about directional influences are needed to infer whether an intervention may work. Theoretical structure may also elucidate optimal intervention points, reducing the need for protracted trial and error in treatment research. Therefore, process models have been developed to consolidate knowledge about chronic pain and inform treatment developments. By aggregating specific knowledge about vulvovaginal pain into an established process model of chronic pain, it becomes reviewable.

The Fear Avoidance Model

While several biopsychosocial models of chronic pain have been developed, e.g. the communal coping model of pain catastrophizing (Sullivan, et al. 2001), the misdirected problem-solving model (Eccleston & Crombez, 2007), and the endurance model (Hasenbring & Verbunt, 2010), none have gained the popularity of the FAM. Accordingly, it has been investigated how well the FAM can be applied to vulvovaginal pain (e.g., Alappattu & Bishop, 2011; Thomtén & Linton, 2013). Both Alappattu and Bishop, and Thomtén and Linton concluded that processes involved in vulvovaginal pain were similar to other pain conditions. Thomtén and Linton also proposed an adaptation of the FAM to accommodate specificities (Figure 1.). The adapted FAM has since found favour among researchers (e.g.,
Flink, Engman, ter Kuile, Thomtén, & Linton, 2017; ter Kuile, Melles, Tuijnman-Raasveld, Groot, & van Lankveld, 2015) and may therefore be a useful model for furthering understanding of vulvovaginal pain.

![Fear Avoidance Model](image)

**Figure 1.** The Fear Avoidance model for PVD (from Thomtén & Linton, 2013. Reproduced with permission of Ekdahl, December 05, 2018).

The popularity of the FAM may be partly explained by its intuitive face validity for both pain sufferers and health care professionals. According to the FAM, pain is sequentially influenced and maintained by emotions, cognitions, behaviours, and biological processes (Vlaeyen & Linton, 2000). Central to this is the circular relationship between the “negative” biopsychosocial factors. This relationship is one directional and implies self-perpetuation. As such, the FAM explains why an episode of acute pain can become chronic. The FAM is also easily adaptable (e.g., Figures 1., and 2.) and has been modified and applied to a variety of conditions, e.g., whiplash (Vangronsveld, Peters, Goossens, Linton, & Vlaeyen, 2007), fibromyalgia (Martínez, Sánchez, Miró, Medina, & Lami, 2011), and tinnitus (Cima, 2018). A paediatric model has been developed to incorporate the maintenance effects of parental pain catastrophizing (Asmundson, Noel, Petter, & Parkerson, 2012). The importance of incorporating context and motivational goals was raised by Vlaeyen, Crombez and Linton...
(2009) in a text where they welcomed evidence-based developments to the model. Thus, the FAM is clearly an applicable and pragmatic model.

The FAM for vulvovaginal pain proposes two sequential relationships, which lead to either continued or resolved pain (Figure 1.). The sufferer perceives pain signals according to prior pain experiences and personal beliefs about pain. The location of pain and intimate context in which pain occurs also play a role in the experience. The interpretation of pain may then lead to two alternatives: (a) an adaptive level of fear, motivating confrontation, and allowing for recovery, or (b) catastrophic thoughts, whereby the maladaptive pain cycle begins. The maladaptive cycle in Thomtén and Linton’s (2013) FAM for vulvovaginal pain has two proposed maintenance pathways (Figure 1.). These pathways may exist separately or co-occur. The primary pathway, present in all variations of the FAM, shows catastrophic thoughts leading to pain related fear and then avoidance. This leads to distress, dysfunction, and disuse. Pain is thus continued, and the cycle begins over. The secondary maladaptive pathway begins identically. However, at fear, the pathway turns toward hypervigilance. This leads to decreased sexual arousal and decreased adaptive physiological responses. Thus, pain is again continued. Thomtén and Linton (2013) also proposed that catastrophizing and fear of relational breakup may motivate endurance of painful sexual activities, which was later supported by Enlund Tuuvas and Lennartsson (2018 [Figure 2.]). This implies that there may be further maladaptive pathways to chronic vulvovaginal pain than those currently proposed.

**Support for Components of the Fear Avoidance Model for Vulvovaginal Pain.**

**Pain stimulus.** An intercourse attempt may cause pain due to inadequate lubrication, muscular tension, sensitivity to hygiene products, or lesions. Pain may also be caused by infections (Arnold, et al., 2006), tampon insertion, or trauma (e.g., Landry & Bergeron, 2011). Sometimes the cause of pain is not apparent.
Figure 2. A suggested alteration to the FAM for PVD accommodating endurance. (From Enlund Tuuvas & Lennartsson, 2018, with permission of Lennartsson, 13th December, 2018).

**Pain perception.** Pain experiences are subjective. Therefore, an universal definition of pain is needed. Pain is currently defined by the International Association for the Study of Pain as: “An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage.” (IASP Task Force on Taxonomy, 1994, p. 209-214). Vulvovaginal pain is often described as cutting or burning (e.g., Bergeron et al., 2001), suggestive of tissue damage, although no damage is present. Some women may have heightened neural sensitivity (e.g., Sutton, Pukall, Wild, Johnsrude, & Chamberlain, 2015) and women with vulvovaginal pain are also more likely than controls to have comorbid pain (Bergeron et al., 2015). This suggests ample previous pain experience. Additionally, Elmerstig (2009) found normative attitudes towards painful sex, suggesting that women with pain may feel conflicted in their perception of pain as a signal of danger. Pain perception in the context of vulvovaginal pain can thus be a complex process.

**Pain catastrophizing.** Pain catastrophizing causes pain to be experienced as more dangerous than is probable (Leung, 2012; Linton, 2013, p. 137). However, pain catastrophizing may also divert attention away from the immediate pain experience (Schütze, Rees, Slater, Smith, & O'Sullivan, 2017). Thus, a feeling of control through “planning for the
worst,” is experienced, without a solution ever being reached. Pain catastrophizing is, therefore reinforced in the short term, but punishing in the long term (Flink, Boersma, & Linton, 2013). The presence of pain catastrophizing has been established in both clinical and nonclinical vulvovaginal pain populations. Catastrophizing was found to be the only unique predictor of intercourse pain intensity in one PVD population (Desrochers, et al., 2009). Further, pain catastrophizing has been seen to influence pain intensity and occurrence of vulvovaginal pain cross-sectionally, and longitudinally (Flink, Engman, ter Kuile et al., 2017), and to significantly mediate solicitous partner responses and pain (Flink, Engman, Thomtén et al., 2017). Pain catastrophizing is thus supported in vulvovaginal pain.

**Fear of pain/partner loss.** Fear of pain has been seen to predict reduced movement in a sample of healthy participants with induced muscle soreness (Trost, France, & Thomas, 2011). Greater fear of pain has also been associated with higher levels of hypervigilance (Crombez, Eccleston, Baeyens, Van Houdenhove, & Van Den Broeck, 1999). Furthermore, Glombiewski et al. (2015) found that people with higher scores on self-report measures of FAM components showed more pronounced physiological fear responses. Women with vulvovaginal pain have also been found to report greater fear of pain than controls (Payne et al., 2007). Moreover, fear of pain has been linked to greater pain sensitivity among PVD sufferers (Desrochers et al. 2009). Fear avoidance beliefs have also been shown to predict future vulvovaginal pain (Ekdahl, Flink, Engman, & Linton, 2018). Additionally, fear of partner loss has been documented (Gordon, Panahian-Jand, McComb, Melegari & Sharp, 2003) and linked with endurance of painful sex (Enlund Tuuvas & Lennartsson, 2018). Fear of pain and/or partner loss are thus supported factors.

**Hypervigilance and reduced arousal.** Hypervigilance entails exaggerated attention to specific stimuli, particularly those associated with fear (e.g., Linton & Flink, 2016, p. 84). Pain intensity and sensitivity has been shown to correlate with hypervigilance (e.g., Herbert
et al. 2014). An emotional Stroop test, where pain was included, supported hypervigilance among women with vulvar vestibulitis (Payne, Binik, Amsel, & Khalifé, 2005). Desrochers, et al. (2008) found that hypervigilance, together with fear and catastrophizing was related to poorer outcomes, and somatically focused anxiety has been found to be more prevalent in a PVD sample than control (Meana & Lykins, 2009). Furthermore, self-monitoring has been linked to reduced arousal (Dove & Wiederman, 2000). There is thus, support for the role of hypervigilance in vulvovaginal pain.

**Avoidance/Endurance.** Avoidance is a learning paradigm which explains reduced frequency of behaviours believed to lead to negative consequences (e.g., Linton, 2013, p. 174-5). Much like pain catastrophizing, avoidance is reinforced in the short term, but punishing in the long term, as it diminished the individual’s behavioural repertoire. The tendency of pain sufferers to avoid activities or movements believed to aggravate symptoms is well documented (e.g., Lundberg, Frennered, Hägg & Styf, 2011). Avoidance of sexual contact is often seen among sufferers of vulvovaginal pain (e.g., Engman, Flink, Thomtén & Linton, 2016). Higher frequency of sex motivated by avoidance of negative relationship consequences has also been documented (Dubé, Bergeron, Muise, Impett, & Rosen, 2017). Furthermore, endurance of painful activities may occur more often in the context of vulvovaginal pain than other pain conditions (Connor, Robinson, & Wieling, 2008). Sufferers of vulvovaginal pain with high avoidance and endurance have been found to have the worst long-term outcomes (Engman et al., 2016). There is, as such, support for both avoidance and endurance in the FAM for vulvovaginal pain.

**Distress/Disuse/Dysfunction.** Pain, chronic or otherwise, leads to distress, and comorbidity with depression is common (e.g., Linton, 2013, Chapter 5). Sufferers of chronic pain have been found to be twice as likely to be have depression as controls (Gureje et al., 2008). Comparable results have been found in relation to vulvovaginal pain (Khandker et al.
Feelings of shame and guilt (e.g., Ayling & Ussher, 2008; Elmerstig, et al., 2008) and reduced partner intimacy have been found in vulvovaginal pain populations (e.g., Connor, et al., 2008; Masheb, Lozano-Blanco, Kohorn, Minkin, & Kerns, 2004). Chronic muscle hypertonicity in the pelvic floor has also been reported (e.g., Goldfinger et al., 2009). Thus, physical dysfunction, reduced intimate contact and emotional distress are seen to co-occur with vulvovaginal pain.

**Partner response.** Thomtén and Linton (2013) proposed that partner responses are the establishing context in which vulvovaginal pain occurs. Research has shown that relationship distress and sexual complaints are common among sufferers of any chronic pain (e.g., Cano, Johansen, Leonard, & de Groot Hanawalt, 2005). Qualitative studies of PVD populations confirm relationship difficulties (Smith & Pukall, 2011). Furthermore, solicitous partner responses have been identified as maintaining vulvovaginal pain (e.g., Rosen, et al., 2013). Solicitous responses have also been seen to correlate with both partners’ pain catastrophizing (Flink, Engman, Thomtén et al., 2017; Rosen et al., 2013). Moreover, partner catastrophizing was found to correlate positively with pain intensity. Higher levels of partner catastrophizing and lower appraisals of women’s self-efficacy have also been seen to correlate with greater pain intensity, independent of women’s own scores (Lemieux, Bergeron, Steben, & Lambert, 2013). Moreover, Benoit-Piau et al. (2018) found that facilitative partner responses buffered the effects of women’s pain catastrophizing on pain intensity. Partner response is therefore a feasible context in which to set the adapted FAM.

Cumulatively, the findings in vulvovaginal pain research show a great deal of support for parts of the FAM. The adapted FAM for vulvovaginal pain is, as such, well-founded.

**Critique of the Fear Avoidance Model**

Despite strong support for the universality of the FAM, there may be reason to reconsider the model, especially in regard to vulvovaginal pain. Pain self-efficacy has been
found to be a better predictor of change in vulvovaginal pain than any part of the FAM (Davis et al., 2015). Additionally, Alappattu and Bishop (2011) stressed that they did not find evidence of the sequential relationships proposed by the FAM. Furthermore, the FAM was formed to explain the development of chronic pain (Vlaeyen & Linton, 2000), not ongoing chronic pain. Most women seeking treatment for vulvovaginal pain however are already chronic sufferers. Additionally, most of the research on the FAM is carried out in WEIRD contexts (Western, Educated, Industrialized, Rich, and Democratic [Henrich, Heine, & Norenzayan, 2010]), using WEIRD samples. The generalizability of the model is thus, limited. Therefore, further investigation of the FAM for vulvovaginal pain is warranted.

The sequential relationships that the original FAM proposes have also been questioned (e.g., Pincus, Smeets, Simmonds & Sullivan, 2010; Ward & Thorn, 2006; Wideman et al. 2013). Ward and Thorn suggest that Cook, Brawer, and Vowles’ (2006) validation of the FAM was flawed, and that pain severity was a better predictor of disability than fear of re-injury, as presented by Cook et al. Additionally, Ward and Thorne suggest a direct pathway between pain catastrophizing and disability. This pathway has since been supported by others (e.g., Flink, Boersma & Linton, 2010; Wideman, Adams, & Sullivan, 2009). Additionally, three alternate pathways through avoidance were proposed by Pincus et al. (2010). However, these avenues do not appear to have been examined further. While research on the FAM has developed since Ward and Thorne’s criticisms, there are few studies which have attempted to validate causal pathways or assess pattern fit.

Furthermore, there is inconclusive evidence about the predictive value of parts of the FAM on disability and chronic pain (e.g., Wertli et al., 2014). While support for the predictive value of catastrophizing is widespread (e.g., Westman, Boersma, Leppert & Linton, 2011) it is not universal (e.g., Lane et al., 2018). In a study which specifically investigated the sequentially of the FAM, Lane et al. found that only work interference and
acute pain intensity were directly associated with pain at six months. Additionally, ter Kuile et al. (2015) illuminated a problem with their findings in support of the pattern fit. Early changes in catastrophizing mediated outcome (pain and function) in an exposure treatment for vaginismus. However, catastrophizing was measured at baseline, six weeks, and completion, whereas pain and function were measured continually. Significant changes were seen in pain and function during the first six weeks. It was conceded by ter Kuile et al. that this disallowed establishment of the chronology of change. Evidently there is mixed support for pattern fit of the FAM, and few studies with sufficient measurements to reliably describe development or return to function.

Weaknesses in the FAM were emphasised by Wideman et al. (2013), who proposed that the model would be better developed into a multidimensional framework with weighted and cumulative risk factors and a variety of pathways. Additionally, Wideman et al. stressed the need for including the influences of micro, meso and macro environments. Moreover, if the linear nature of the FAM is speculative, it cannot be considered to illustrate causality. Thus, critical evaluation of interventions informed by the model is necessary. Furthermore, as the FAM lacks alternative rehabilitation pathways, there is a blind-spot in the development of treatments for chronic pain. With empirically based refinements, researchers and clinicians may be better equipped to identify treatment hinders and maximal intervention disposition. Alternatively, growing unsupportive evidence may motivate a paradigm shift away from the FAM. In summary, there is need for further examination of the pattern of change in the fear avoidance model, especially in relation to treatment interventions.

**Aim**

The primary aim of this thesis was to examine the feasibility of the proposed CBT+ group treatment for provoked vulvodynia. Feasibility entails treatment gains in pain
intensity, sexual function, and partner response. Furthermore, levels of compliance, completion, perceived credibility, and change in depression and anxiety will be assessed.

The secondary aim of this study was to examine the process of change seen in integral components of the FAM: catastrophizing, fear, and avoidance. Thus, return to function and pattern fit between observed processes and those proposed by the FAM will be examined.

**Hypotheses**

- The treatment will reduce women’s’ pain.
- The treatment will increase women’s sexual function.
- The treatment will reduce the frequency of solicitous and punishing partner responses and increase the frequency of facilitative partner responses.
- The treatment will reduce symptoms of depression and anxiety.
- The pattern of change seen in catastrophizing, fear, and avoidance, will fit the pattern proposed by the fear-avoidance model. Changes in catastrophizing are expected to precede changes in fear, which are expected to precede changes in avoidance.

**Method**

The data presented in this thesis is a subsample of that collected in the single case arm of the VENUS study. The VENUS study is a treatment trial of group CBT with partner sessions for PVD. There are two trial arms, (a) an RCT, and (b) a SCED with repetitions. The study is a collaboration between Örebro, Maastricht, and Leiden Universities. All procedures occurred in a city in central Sweden with >100,000 inhabitants.

**Design**

This study used a replicated single case, quasi-experimental AB design. As this study had multiple participants, but no direct comparison between participants, it was considered a replicated design rather than a multiple baseline design. The main aims of a SCED are to
clarify whether change is evident between phases, and whether that change can be reliably attributed to the experimental manipulation. Thus, repeated measures were taken of the dependent process variables for the duration of this study. The baselines were either nine or six weeks long and acted as the control to the 16 week treatment phase. Outcome measures were also taken at inclusion, post, and FU. Idiosyncratic target measures were not used as this thesis also aimed to examine change and pattern fit to the FAM, not exclusively individual symptom relief. Practical constraints disallowed randomization of baselines in this study. Ergo, it must be considered quasi-experimental. Conclusions about causality are thus not recommended by all (e.g., Onghena & Edgington, 2005). However, quasi-experimental studies with strong designs may provide valid results (e.g., Shadish, Cook & Campbell, 2002, p. 486). Thus, with acknowledgement of validity threats, causal conclusions may be drawn.

Visual analysis has been the analysis method of choice for SCED (e.g. Morley, 2017, p. 87). The use of repeated measures leads to autocorrelation, and data is likely to be non-parametric, limiting the choice of statistical tests. However, Morley (2017, p. 151) proposes that thorough analysis should be a synthesis of both visual and statistical methods. Statistical analyses are more replicable than visual analyses, and the risk of Type II errors is reduced as subtle changes in data may be detected. Previous treatment studies for vulvovaginal pain have shown mixed effects (e.g., Corsini-Munt et al., 2014; ter Kuile, et al, 2006) suggesting that the results of this study would not be clear-cut. Additionally, the study was believed to be unique in its aim to assess pattern fit of the FAM using SCED. Ease of replication was therefore desirable. Thus, both visual and statistical analyses were motivated.

**Participants and Procedures**

**Selection criteria.** The participants were cis-women (18-45 years) and their male partners. All couples were sexually active (≥ 3 months) with PVD symptoms during ≥75% of intercourse attempts ≥ 6 months. No clinical diagnosis was necessary but medical
examination was encouraged. Exclusion criteria were; partners unwilling to participate, pregnancy or birth < 1 year, ongoing genital tissue damage or infection, sexual inactivity > 1 month, and ongoing psychological trauma or serious mental disorders (women only).

**Recruitment procedure.** Recruitment for both arms of the VENUS study occurred consecutively, from October 2017 to January 2018. Recruitment for the single case continued after recruitment to the RCT was complete. Recruitment ended 6 weeks before treatment start to allow for baseline. Posters advertising the RCT were displayed at local health care clinics and public buildings. Adverts were displayed on the Örebro University homepage and social media, and the study was featured on local radio and television. Prospective participants registered interest by email and written background information was sent to all.

Initial telephone screening was held by a licensed psychologist and consisted of structured interviews with all women. Eligible couples were invited to a second interview, held by a senior year psychologist student. Informed consent was collected from both partners. If grounds for exclusion were not forthcoming in the structured clinical interview, the Mini International Neuropsychiatric Interview 6.0 (MINI) (Lecrubier et al., 1997) was conducted with the women. If no grounds for exclusion were found, the couple was informed of their eligibility. Couples wishing to participate completed initial measurements on site.

**Participants.** The women were Swedish or Finnish and their educational level was secondary or tertiary. One female participant had children. Male partners were Swedish or Finnish and had primary, secondary or vocational educations. All women had previously sought care for their symptoms, but only one had received a diagnosis (Vestibulitis). Three had received treatment (CBT, medicated cream, or physiotherapy). One woman stated that she was currently receiving care (physiotherapy). Two participants reported suffering from other chronic pain. Data on other pain problems was missing for the remaining participants.
Participant flow.

Figure 3. Recruitment and attrition.

Measures

All measures were digital, in Swedish, and answered based on experiences during the past four weeks. Partner measures were taken at inclusion, post treatment, and at FU. Women’s measures were taken weekly from inclusion until post treatment, and at FU.

Demographic information. The demographic variables for both women and their partners were; age, nationality, parental status, and level of education. The women were also asked about their pain debut, the length of their complaint, whether their pain was primary or secondary, if they had previously sought care, whether they also suffered from other pain, and/or had given birth.

Primary outcome measures. Primary outcome measures were taken weekly, and at inclusion, post treatment, and FU. Weekly measures of pain and function were presented as z-scores in figures with the combined catastrophizing, fear, and avoidance (c-f-a) process variable. Inclusion, post treatment, and FU measures were presented in tables.

Vulvovaginal Pain. A two-item numerical rating scale (NRS) with a total score ranging from 0-20 was used to measure the a) intensity and b) unpleasantness of experienced vulvovaginal intercourse pain. Item scores of 0 represented no symptoms, and 10, unbearable symptoms. Numerical scales of pain have been shown to be sensitive to change, have good reliability (ICC = .95 (95% CI = .93–.96)), and validity (correlation between Visual Analogue Scale for pain rating and NRS, $r = .941$, $p < .001$ in a sample with knee pain) (Alghadir,
Anwer, Iqbal, & Iqbal. 2018). According to Dworkin et al., (2008) the recommended level for meaningful change for NRS in clinical trials are reductions of 30%.

**Sexual Function.** The full 19 item version of the Female Sexual Function Index (FSFI) (Bayer AG, Zonagen, Inc. & Target Health Inc., 2018) was used to assess sexual function at inclusion, post treatment and FU. A 6 item version was used in weekly measures. The FSFI assesses six domains of sexual function (desire, arousal, lubrication, orgasm, satisfaction, pain) and high scores represent good sexual function. For the 6-item scale the minimum score was 1, and maximum 12. For the full scale the minimum was 2, and maximum 36. Items are answered using Likert-type scales of various lengths. Thus, each domain score is multiplied by a factor before being summed into the full-scale score (Bayer AG, Zonagen, Inc. & Target Health Inc., 2018.). The FSFI has been shown to have good test-retest reliability and internal consistency (full scale $r = .88$, Cronbach’s $\alpha = 0.97$, Rosen et al. 2000). The FSFI has also demonstrated discrimination between respondents belonging to control, vulvodynia, and Female Sexual Arousal Disorder groups (Rosen et al., 2000; Masheb, Lozano-Blanco, Kohorn, Minkin, & Kerns, 2004).

**Partner response.** Partner responses were measured using an adapted version of “The partner responses questionnaire to PVD” (MPI-SR [Rosen, Bergeron, Sadikaj, & Delisle, 2015]). The MPI-SR is based on the West Haven-Yale Multidimensional Pain Inventory Significant Other Response Scale (Kerns, Turk, & Rudy, 1985) and was developed specifically for pain in sexual situations. Accordingly, there are two versions, one for the person in pain, and one for their responding partners. Questions are asked on two subscales solicitous (6 items) and punishing (4 items). The reliability of the MPI-SR was deemed to be acceptable to high (solicitous subscale, $r = .72$ and $r = .85$, punishing subscale, $r = .73$ and $r = .74$, for women and partners respectively (Rosen et al. 2015). An expanded version of the MPI-SR which included a facilitative subscale, was used in this study (Rosen et al. 2015).
Questions regarded the frequency of specific responses and were answered using a 6 point Likert-type scale. Scores of 1 represented “Never” and 6 “Very often”. The minimum score for each domain was 6 (4 for punishing) and the maximum score for each domain was 36 points (24 for punishing). No norms were found for the measure, thus, the criteria for change was set at $> \pm 10\%$.

**Secondary outcome measures.** Secondary outcome measures were taken at inclusion and/or post treatment and FU. They were presented in tables and text.

**Compliance, completion, and perceived credibility.** Compliance was defined as the percentage of sessions attended by each woman/couple. Completion was assessed using attrition rates. Perceived credibility was assessed at post treatment and FU using the following questions: (a) “To what degree do you consider that your (partner’s) intercourse pain has been alleviated by the treatment you have received during this study?” answered on a 6-point scale ranging from “complete remission” to “increased pain”. (b) “To what degree do you consider that the general quality of your sexlife has been improved by the treatment you received during this study?” This was answered on a 6-point scale ranging from “absolute improvement (never been better)” to “deterioration”. (c) “On a scale of 0 to 10, how satisfied are you, on the whole, with the treatment you have received.” Zero represented “very dissatisfied” and 10, “very satisfied”.

**Depression and anxiety.** The following measures were taken by female participants at inclusion, post treatment and FU.

**PHQ-9.** The PHQ-9 is a nine item self-report scale used for screening of depressive disorders. The scale was developed as a subscale of the PHQ (Patient Health Questionnaire) and subsequently validated for standalone use (Kroenke, Spitzer, & Williams, 2001). All items are answered on a 4 point scale ranging from 0-3. Total scale scores range from 0-27, with higher scores representing more serious symptoms. The internal consistency of the
English PHQ-9 is considered excellent (Cronbach’s $\alpha = 0.86$ in an obstetrics/gynaecology sample) (Kroenke, et al., 2001). The test-retest reliability (48 hours) was also found to be excellent ($r = .84$). Cut-off values were: 0-4 (minimal symptoms), 5-9 (mild symptoms), 10-14 (moderate symptoms), 15-19 (moderately severe symptoms) and 20-27 (severe symptoms) (Kroenke, et al., 2001).

**GAD-7.** The GAD-7 is a 7 item questionnaire developed to assess Generalised Anxiety Disorder symptoms (Spitzer, Kroenke, Williams, & Löwe, 2006). All items are answered on 4 point scales ranging from 0-3. Total scale scores range from 0-21, with higher scores representing serious symptoms. The GAD-7 was found to have good test-retest reliability ($r = .83$) and excellent internal consistency (Cronbach’s $\alpha = 0.92$) in a large U. S. primary care sample (Jordan, Shedden-Mora, Lowe, & Loewe, 2017). Cut-off values were: $\leq$ 4 (negligible symptoms), 5-9 (mild), 10-14 (moderate), and 15-21 (severe symptoms of generalised anxiety) (Spitzer et al., 2006).

**Process measures.** Individual items were taken from scales in the women’s weekly measures to represent the process variables. As items had differing scales, $z$-scores were calculated. All items were positively skewed. Process measures were presented in figures.

**Pain catastrophizing.** According to Thomtén and Linton’s (2013) FAM, pain catastrophizing is cognition about the consequences of pain. Consequently, the following item from Klaasen and ter Kuile’s (2009) Vaginal Pain Cognitions Questionnaire (VPCQ) was used to represent pain catastrophizing. “I have the following thoughts about vaginal penetration: I am afraid that I can not do anything to change the pain from penetration” (…: Jag är rädd att inte kunna göra något för att förändra penetrationssmärtan”). The question was answered on a 7 point Likert-type scale. Scores of 0 represented “Not at all applicable”, and 6 “Very strongly applicable”. The reliability of the Catastrophic and pain cognitions subscale of the VPCQ was found to be good ($r = .86$). The internal consistency was found to
be excellent \( (r = .86, \text{Cronbach’s } \alpha = 0.91) \) (Klaasen & ter Kuile, 2009). The item had a factor loading of .789 (EFA).

**Fear of pain.** Linton (2013, p. 101) describes fear as a primarily physiological and present focused anxiety response to, or in anticipation of, a frightening stimulus. Thus, the following item was chosen from the Catastrophic and pain cognitions subscale of the VPCQ to measure fear of pain. “I have the following thoughts about vaginal penetration: I am afraid of cramping up during penetration” (…: Jag är rädd att spänna mig under penetration). The question was answered on a 7 point Likert-type scale. Scores of 0 represented “Not at all applicable” and 6 “Very strongly applicable”. The item had a factor loading of .724 (EFA) (Klaasen & ter Kuile, 2009).

**Avoidance.** Avoidance is the act of disengaging from an experience, whether overtly, or covertly, as the experience is deemed too unpleasant to engage in at that point in time (e.g., Linton & Flink, 2016, p. 103). The first item of the Penetration Behaviours Questionnaire (PBQ) (ter Kuile et al., 2007) was used to measure avoidance as this item was considered least likely to be contaminated by home practices in the treatment. The question read “Have you had intercourse with complete vaginal penetration of your partner’s penis during the past four weeks?” (Har du haft samlag med fullständig vaginal penetration av din partner’s penis under de senaste fyra veckorna?). Possible answers were “No attempts made”, “Have attempted but not succeeded”, “Have been successful on some attempts”, and “Have been successful on all attempts”. The internal consistency of the PBQ was found to be acceptable (Cronbach’s \( \alpha = 0.72 \) to 0.79) (ter Kuile et al., 2007).

**Combined catastrophizing, fear, and avoidance (c-f-a).** This variable was presented together with weekly measures of pain and function to illustrate how changes in process variables covaried with changes in the outcome variables. The variable was created by summing the z-scores from the catastrophizing, fear, and avoidance items and dividing by 3.
Other measures taken but not included in analysis. See Appendix.

Experimental Manipulation

Baseline. The treatment start date was set during recruitment. Thus, baselines lengths were dependent on the date of secondary screening interviews and were either nine or six weeks.

Table 1
An outline of the treatment program, with select content and practice.

<table>
<thead>
<tr>
<th>Week</th>
<th>Session</th>
<th>Content</th>
<th>Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>C1</td>
<td>Treatment expectations</td>
<td>‘Taking care of the vulva’</td>
</tr>
<tr>
<td>3</td>
<td>G1</td>
<td>Treatment expectations and FAM</td>
<td>Formulation of treatment goals</td>
</tr>
<tr>
<td>4</td>
<td>G2</td>
<td>Relationship between fear and muscle tension</td>
<td>Progressive relaxation exercises</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Looking at your vulva</td>
</tr>
<tr>
<td>5</td>
<td>G3</td>
<td>Anatomy of the pelvic floor; Fear and pelvic floor muscle tension; introduction of graded exposure</td>
<td>Pelvic floor and breathing exercises; Graded exposure</td>
</tr>
<tr>
<td>6</td>
<td>G4</td>
<td>Thoughts, feelings, and behaviour</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>G5</td>
<td>Sexual arousal and pain</td>
<td>Body scan and sensate focus (not genitals), with partner</td>
</tr>
<tr>
<td>8-9</td>
<td>C2</td>
<td>Progress check and problem solving</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>G6</td>
<td>Woman and partner responses to pain</td>
<td>Sensate focus exercise (with genitals), with partner</td>
</tr>
<tr>
<td>12</td>
<td>G7</td>
<td>Communication</td>
<td>(re)Introduction of coitus Communication about sex</td>
</tr>
<tr>
<td>14</td>
<td>G8</td>
<td>Negotiation</td>
<td>Negotiation about sex</td>
</tr>
<tr>
<td>16</td>
<td>G9</td>
<td>Tools for the future</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>G10</td>
<td>Evaluation of the treatment</td>
<td>(Post treatment measures)</td>
</tr>
<tr>
<td>19-22</td>
<td>C3</td>
<td>Progress check and problem solving</td>
<td></td>
</tr>
<tr>
<td>25-26</td>
<td></td>
<td>Assessment</td>
<td></td>
</tr>
<tr>
<td>+ 3 M</td>
<td></td>
<td>Follow-up</td>
<td>(Follow-up measures)</td>
</tr>
</tbody>
</table>

Note. C = Couples’ session, G = Group Session
**Treatment.** The CBT treatment consisted of three couples’ sessions, and ten group sessions for the women. All sessions were led by two licensed psychologists (see Table 1 for an overview). The couples’ sessions were scheduled to take 45-60 minutes and took place (a) prior to the first group session, (b) after the first five group sessions, and (c) after completion of all group sessions. At these sessions couples were able to discuss their personal situation. Focus lay on expectations and ensuring that treatment practices were being executed. Problem-solving skills were discussed where needed. The group sessions were initially held weekly, and then at 2 week intervals after the second couples’ session. Group sessions were scheduled to take two hours, and each session had a specific theme. Sessions were based on the protocol by ter Kuile and Weijenborg, (2006). Home practice and/or reading material were given at most sessions, but no record of completion was kept. All materials were available digitally throughout treatment.

**Assessment and Follow-up.** Telephone assessments were held with all women circa one month after treatment completion. The FU session was a women’s group session.

**Ethical considerations.** The Department of Psychology at Örebro University approved this study per ethical guidelines for master’s theses. The VENUS study was approved by the regional board of ethics for scientific studies in Uppsala [dnr2017-289]. There was no evidence of interdependence between participants, or between participants and others involved in the study. The author of this thesis had no contact with participants.

**Analyses**

**Data processing.** All preliminary data analysis and calculation of z-scores was carried out in IBM SPSS Statistics for Windows (Version 25.0., 2017). Z-scores were calculated using pooled participant data, separately for each phase. Graphs and PEM were calculated in Microsoft Excel (2016). Calculations of Fisher’s exact probability were done using the College of St Benedict St John’s University (2018) Exact r x c contingency table.
**Visual analysis.** Visual analysis entails judgement of graphically presented data. One or more analyst assesses whether there is evidence of change, and whether observed changes are meaningful. Key to interpretation is that graphics are clear, precise, and equivalent for all participants (Morley, 2017, p. 89). Visual analysis should be done systematically, using criteria for change set prior to analysis. Systematic analysis often includes: analysis of the central location of each phase, evidence of trend/s within each phase, changes in level, latency of change between phases, variability within phases, and degree of overlapping data points in each phase (e.g., Kazdin, 2011, p. 28; Morley, 2017, p95). The magnitude of the following changes were noted; median: differences of ≥ 1 SD, variability: directional changes ≥ 1SD on more than one occasion, level: directional changes of >1SD at phase change. No criteria were set for trend or latency.

**Statistical analysis.** The statistical test considered most appropriate for this study was *percentage exceeding the median of the baseline* (PEM). PEM uses the median of the baseline as a reference point to which all data in the intervention phase is compared. The number of intervention phase scores under, or over, the median (depending on the directionality of the research hypothesis) are counted (e.g., Ma, 2006). PEM is then calculated using the formula: 

\[
\text{PEM} = \frac{\text{number of data points in treatment phase exceeding the median of the baseline}}{\text{total number of data points in the treatment phase}}
\]

and is expressed as a value between 0 and 1.

PEM should be considered a measure of effect size (Scruggs & Mastropieri, 1998) with scores of ≥ .9 denoting highly effective, ≥ .7-.89 moderately effective, and ≥ .5-.69 interventions with little effect. Scores of < .5 are considered to represent ineffective treatments. PEM cannot detect stability, trend, or the magnitude of change (Scruggs & Mastropieri, 1998) and can only be seen as supporting evidence of difference between phases. Fisher’s exact probability test was used to calculate probability values as the sample
size was presumed to be too small to use Mood’s median test accurately. The significance level was set at $p < .05$.

To fulfil the second aim of this thesis, a customized “PEM-count” (PEM4) was developed. Periods of 4 weeks were chosen, as the length of the treatment was divisible by 4, and gradual, differentiated change in the process variables was of interest. The median was found for the baseline and applied to the first 4 weeks of treatment. PEM for this period was then calculated. The median of that 4 week treatment period was then applied to the following 4 weeks of treatment, and so on, until treatment completion. This was replicated for the variables: catastrophizing, fear and avoidance. PEM4 was then displayed in a figure (denoted $c$), without being superimposed over the original data points and used for visual analysis of pattern fit. Criteria for change were not set as no benchmark was found.

**Results**

**Intervention fidelity.** At the time of writing this thesis, the videos of group treatment sessions had not been reviewed. It was therefore not possible to assess intervention fidelity.

**Results by Couple**

**Couple 1.**

**Primary outcome measures.** Pain and c-f-a displayed downward trends in the baseline, but upward trends in the treatment phase for woman 1 (Figure 3., a). Function displayed no trends throughout (Figure 3., a). Directional change in pain and c-f-a occured in the first week of treatment. Pain, function and c-f-a displayed stability throughout, but there are signs of instability in pain in the treatment phase. Changes in medians over 1 SD and changes in level between the baseline and treatment phases were not detected in pain, function or c-f-a for woman 1. There was a varying degree of overlap between data from baseline and treatment (pain 56.25 %, function 93.33 %, c-f-a 67.75 % [Figure 3., a]). It was concluded that weekly measures of pain, function and c-f-a for woman 1 indicate deterioration during treatment.
Figure 3. Weekly measures of pain, function, and combined catastrophizing, fear, and avoidance; and weekly measures of catastrophizing, fear, and avoidance, for woman 1.

Dashed lines indicated the median of the previous phase. Decimal scores denote percentage improvement compared to the median of previous phase (PEM). Fisher’s exact test is reported where $p \leq .05$. Black phase break denotes baseline to treatment. Red phase break denotes end of treatment.
PEM pain (a, .06, p = .02) and c-f-a (a, .19, p = .004) support significant deterioration during treatment. PEM function (.6, small effect size [Figure 3., a]) indicated negligible change in treatment. The trends seen in weekly measures reflect woman 1’s inclusion and post measurements of pain but not function (Table 2). Woman 1 reported a 25% total increase in pain and a 21.47% total decrease in function from inclusion to FU, with greatest deterioration occurring during treatment.

Solicitous responses were reported to increase by both partners from inclusion to post treatment (woman, 13.33%, man, 30%). Increases were lost for woman 1 and partially sustained for man 1 at FU. Punishing responses were absent throughout. Facilitative responses were reported heterogeneously. Woman 1 reported a drop from high to low frequency of facilitative responses during treatment (-46.67%). All losses were regained at FU. Man 1 reported stability.

**Secondary outcome measures.** Couple 1 had an excellent level of treatment compliance. Woman 1 considered the treatment to have alleviated her pain a little at both post treatment and FU. This is at odds with her NRS scores (Table 2) as increases in pain were reported. Man 1 considered her to have had no improvement in pain at post, and deterioration at FU. Woman 1 considered there to be a large improvement un the quality of her sex life at both post and FU. This is also at odds with her FSFI scores (Table 2) where deteriorations were seen. Man 1 considered there to be no improvement in the quality of his sex life at post, and a deterioration at FU. Woman 1 reported no symptoms of depression or anxiety at inclusion or post treatment but did report negligible symptoms at FU.

**Process measures.** Catastrophizing displayed a steep downward trend in the baseline, but an upward trend in the treatment phase. Directional change occurred immediately after phase change. Fear and avoidance displayed no trends in the baseline. Fear displayed a slight upward
A trend in the treatment phase beginning in the eighth weeks of treatment (week 17). Trend in avoidance was unclear in the treatment phase (Figure 3., b). Catastrophizing, fear and avoidance displayed stability throughout, but there are signs of instability in avoidance in the treatment phase. Changes in medians over 1 SD were not detected in catastrophizing, fear or avoidance for woman 1. There was a large degree of overlap between data from baseline and treatment for catastrophizing (100%). Less overlap was seen in fear (37.5%) and no overlap was seen in avoidance (0% [Figure 3., b]). It was concluded that weekly measures of catastrophizing, fear and avoidance for woman 1 indicate deterioration during treatment.

Table 2

Scale scores and percentage change for woman 1 (24, primary pain, 7 years) and partner (22)\(^a\)

<table>
<thead>
<tr>
<th></th>
<th>NRS</th>
<th>FSFI</th>
<th>MPI_S</th>
<th>MPI_P</th>
<th>MPI_F</th>
<th>PHQ9</th>
<th>GAD7</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion</td>
<td>6</td>
<td>24.50</td>
<td>24</td>
<td>4</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>9(15)</td>
<td>19.90(-13.53)</td>
<td>28(13.33)</td>
<td>4(0)</td>
<td>13(-46.67)</td>
<td>0 (0)</td>
<td>0(0)</td>
<td>6/10</td>
</tr>
<tr>
<td>Follow-up</td>
<td>11(10)</td>
<td>17.20(-7.94)</td>
<td>23(-16.66)</td>
<td>4(0)</td>
<td>27(46.67)</td>
<td>2(7.41)</td>
<td>2(9.52)</td>
<td>7/10</td>
</tr>
<tr>
<td>Total(^b)</td>
<td>25%</td>
<td>-21.47%</td>
<td>-3.33%</td>
<td>0%</td>
<td>0%</td>
<td>7.41%</td>
<td>9.52%</td>
<td></td>
</tr>
</tbody>
</table>

Partner

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion</td>
<td>17</td>
<td>4</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>26(30)</td>
<td>4(0)</td>
<td>29(6.67)</td>
<td></td>
<td></td>
<td>5/10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow-up</td>
<td>24(-6.67)</td>
<td>4(0)</td>
<td>28(-3.33)</td>
<td></td>
<td></td>
<td>10/10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total(^b)</td>
<td>23.33%</td>
<td>0%</td>
<td>3.33%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(\text{Note.}\) NRS = Numerical Rating Scale for pain, FSFI = Female Sexual Function Index, MPI S = MPI-SR solicitousness scale. MPI P = MPI-SR punishing scale. MPI F = MPI-SR facilitating scale. TS = Treatment Satisfaction. Scores in brackets = % change since previous measure occasion in relation to full scale score.

\(^a\)Compliance: group, 90% (did not attend session 5, but participated in a compensatory telephone session), partner sessions, 100%.

\(^b\) Change in relation to full scale score.
PEM catastrophizing (0.06, \( p = .003 \)), fear (0.25, \( p < .001 \)) and avoidance (0.44, \( p < .001 \)) all indicated significant deterioration during treatment (Figure 3., b). PEM4 for catastrophizing, fear and avoidance indicated greatest deterioration in the middle 8 weeks of treatment (Figure 3., c). Increases in catastrophizing were seen to occur before increases in avoidance. Decreases in avoidance were not seen to lag but occurred simultaneously or in advance of decreases in catastrophizing. PEM4 indicated a degree of covariance between catastrophizing and avoidance. The same was not seen in fear, suggesting that it followed a different process tangent for woman 1.

Couple 2.

**Primary outcome measures.** Pain displayed no trend in the baseline but a slight downward curvilinear trend in the treatment phase (Figure 4., a). Change in direction occurred in the third week of treatment (week 12). Function displayed no trend in either phase. C-f-a displayed slight downward trends in the baseline and downward trend in the treatment phase (Figure 3., a). The latency of change for c-f-a was unclear. Pain, function and c-f-a displayed stability throughout, but there are indications of instability in pain and function in the treatment phase. Changes in medians over 1 SD and changes in level between baseline and treatment were not detected in pain, function or c-f-a for woman 2. There was a little overlap between baseline and treatment phase data for pain and c-f-a (pain 37.5 %, c-f-a 18.75 %), but a large degree of overlap in function (68.75 % [Figure 4., a]). It was concluded that weekly measures of pain and function for woman 2 indicate small improvements or negligible change during treatment, while c-f-a displays some improvement.

PEM pain (.63, small effect size, \( p = .01 \)) and function (.88, moderate effect size, \( p < .001 \)) support improvements during treatment. PEM c-f-a (1, large effect size, \( p < 0.001 \)) supports significant improvement during treatment (Figure 4., a). The trends seen in weekly
**Figure 4.** Weekly measures of pain, function, and combined catastrophizing, fear, and avoidance; and weekly measures of catastrophizing, fear, and avoidance, for woman 2.

Dashed lines indicated the median of the previous phase. Decimal scores denote percentage improvement compared to the median of previous phase (PEM). Fisher’s exact test is reported where $p \leq .05$. Black phase break denotes baseline to treatment. Red phase break denotes end of treatment.
measures reflect woman 2’s inclusion and post measurements of pain and function (Table 2). Woman 2 reported negligible change in pain during treatment, but a 60% total decrease in pain from inclusion to FU. Woman 2 reported a 19.11% increase in function during treatment, but this had halved at FU (Table 3).

**Secondary outcome measures.** Couple 2 had an excellent level of treatment compliance. Woman 2 considered the treatment to have alleviated her pain a lot at both post treatment and FU. This is supported only by her NRS scores at FU (Table 3). Her partner considered her to have had moderate improvement in pain at post, and no improvement at FU. Woman 2 considered there to be a moderate improvement in the quality of her sex life at post and a small improvement at FU. This is supported by her FSFI scores (Table 3). Man 2 considered there to be small improvements in his sex life at both post and FU. Woman 2 reported mild to moderate levels of depression and anxiety at inclusion. At post-treatment both depression and anxiety were reported to have dropped into the mild range. However, by FU levels of both depression and anxiety had increased above inclusion levels.

Solicitous responses were only reported to have increased by woman 2 during treatment (23.33%). Punishing responses were also reported heterogeneously. Woman 2 reported a 30% decrease during treatment which was sustained at FU. Man 2 reported stability. Facilitative responses were unanimously reported to have increased at post (woman 26.67%, man 16.67%) with gains partially sustained at FU. Woman 2 consistently reported higher facilitative responses than her partner.

**Process measures.** Catastrophising displayed a downward trend in the baseline but no trend in the treatment phase. Change occurred in the fifth week of treatment. Fear and avoidance displayed no trends throughout, however, there are indications of an emerging downward trend in the final weeks of the treatment phase. Catastrophizing and avoidance displayed stability throughout. Fear displayed stability in the baseline, but instability (> 1SD)
in the final weeks of the treatment phase. Changes in median >1 SD were not detected in catastrophizing, fear or avoidance. Small changes in level were observed in catastrophising and avoidance, indicating improvement but missing data at week 9 disallowed definitive judgements. A large degree of overlap between data from baseline and treatment was seen in catastrophizing (100%) and avoidance (87.5%). No overlap between phases observed in avoidance (see Figure 4, b). It was concluded that weekly measures of process variables for woman 2 show improvement in catastrophising and avoidance, but no improvements in fear. There are however indications of change in the last 5 weeks of treatment.

Table 3

*Scale scores and percentage change for woman 2 (34, secondary pain, 1 year) and partner (31)*

<table>
<thead>
<tr>
<th></th>
<th>NRS</th>
<th>FSFI</th>
<th>MPI_S</th>
<th>MPI_P</th>
<th>MPI_F</th>
<th>PHQ9</th>
<th>GAD7</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion</td>
<td>18</td>
<td>3.50</td>
<td>23</td>
<td>10</td>
<td>28</td>
<td>10</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>19(50)</td>
<td>10(19.11)</td>
<td>30(23.33)</td>
<td>4(-30)</td>
<td>36(26.67)</td>
<td>6(-14.82)</td>
<td>7(-9.53)</td>
<td>10/10</td>
</tr>
<tr>
<td>Follow-up</td>
<td>6(-65)</td>
<td>7(-8.82)</td>
<td>29(-3.33)</td>
<td>4(0)</td>
<td>29(-23.33)</td>
<td>12(22.22)</td>
<td>18(53.38)</td>
<td>10/10</td>
</tr>
<tr>
<td>Totalᵇ</td>
<td>-60%</td>
<td>10.29%</td>
<td>20%</td>
<td>-30%</td>
<td>3.33%</td>
<td>7.41%</td>
<td>47.62%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Partner*

|            |      |      |       |       |       |      |      |    |
| Inclusion  | 22  | 8    | 24    |       |       |      |      |    |
| Post       | 23(3.33) | 8(0) | 29(16.67) |       | 7/10 |
| Follow-up  | 25(6.67) | 8(0) | 27(-6.67) |       | 8/10 |
| Totalᵇ     | 10% | 0%   | 10%   |       |       |      |      |    |

*Note.* NRS = Numerical Rating Scale for pain, FSFI = Female Sexual Function Index, MPI_S = MPI-SR solicitousness scale. MPI_P = MPI-SR punishing scale. MPI_F = MPI-SR facilitating scale. TS = Treatment Satisfaction. Scores in brackets = % change since previous measure occasion in relation to full scale score.

ᵇCompliance, group, 90% (did not attend session 5), partner sessions, 100%.

ᵇin relation to full scale score.
PEM catastrophizing (.94, p < .001) and avoidance (1, p < .001) supported significant improvement during treatment (large effect sizes). PEM fear (1, p < .001) also indicated significant improvements, but the magnitude of change was negligible, and therefore discounted (Figure 4., b). PEM4 (Figure 4., c) indicated a degree of covariation in fear and avoidance. The same was not seen in catastrophizing, suggesting it followed a different process tangent for woman 2.

**Couple 3.**

**Primary outcome measures.** Pain displayed no trend in the baseline but an upward trend in the treatment phase for woman 3. Directional change occurred in the last week of the baseline. However, increases sharply dropped in the thidr week of treatment, whereby a new, gentler upward trend began. Function displayed an upward trend in the baseline, but no trend in the treatment phase. Directional change occurred at phase break. C-f-a displayed a slight downward trend in the baseline and an upward trend in the treatment phase (Figure 5., a). Directional change followed the same pattern as pain. Pain and c-f-a displayed stability throughout. Function displayed stability in the baseline, but instability (> 1SD) in the treatment phase. Changes in medians over 1 SD were not detected in pain, function of c-f-a for woman 3. Changes in level between baseline and the treatment phase were observed in pain, function and c-f-a, but these were not sustained in any variable. There was a large degree of overlap data from baseline and treatment phases for all three variables (pain 87.5 %, function 87.5 %, and c-f-a 87.5 % [Figure 5., a]). It was concluded that weekly measures of pain, function and c-f-a for woman 3 indicate deterioration during treatment.

PEM pain (0, p = .002,), function (.38) and c-f-a (0.13) support deterioration during treatment (Figure 5., a). The trends seen in weekly measures reflect do not reflect woman 3’s inclusion and post measurements of pain as decrease were reported at both post and FU (total
Figure 5. Weekly measures of pain, function, and combined catastrophizing, fear, and avoidance; and weekly measures of catastrophizing, fear, and avoidance, for woman 3. Dashed lines indicated the median of the previous phase. Decimal scores denote percentage improvement compared to the median of previous phase (PEM). Fisher’s exact test is reported where \( p \leq .05 \). Black phase break denotes baseline to treatment. Red phase break denotes end of treatment.
30 %). Overall, weekly measures of function do reflect those seen in Table 4 (29.7% total increase) but it was observed that improvements occurred largely during baseline.

Solicitous responses were reported heterogeneously. Woman 3 reported small fluctuations (up to 10%) in solicitous responses. Man 3 reported a 33.33% increase during treatment, but FU data is missing. Punishing responses were uniformly reported to be very low or absent, however, man 3 reported a 10% increase from post to FU. Facilitative responses are uniformly reported to be stable and high or very high. Woman 3 reported stability while man 3 reported a slight increase in facilitative responses (10% total increase).

**Secondary outcome measures.** Couple 3 had an excellent level of treatment compliance. Woman 3 considered the treatment to have alleviated her pain moderately at post treatment and a lot at FU. This is supported by her NRS scores (Table 4). Her partner considered her to have had small improvements in pain at post and moderate improvements at FU. Woman 3 considered there to be a large improvement in the quality of her sex life at both post treatment and FU. This is also supported by her FSFI scores (Table 4). Man 3 considered there to be small improvements in his sex life at both post and FU. Woman 3 reported negligible symptoms of depression and mild anxiety at inclusion. Symptoms had increased at post-treatment. At follow-up symptoms had receded again.

**Process measures.** Downward trend was observed in catastrophizing in the baseline phase. Catastrophizing displayed an upward trend in the treatment phase and change in direction occurred at phase break. Avoidance displayed no clear trend in either phase. Fear displayed no trend in the baseline but an upward trend in the treatment phase. Changes in median > 1 SD were not found in any variable, but large changes in fear were observed (0.89 SD [Figure 5., b]). Changes in level >1SD between baseline and treatment were detected in catastrophizing (1.28 SD) and avoidance (1.05 SD). Changes in level were not sustained. Stability was unanimously detected in fear throughout. Signs of instability
Table 4

Scale scores and percentage change for woman 3 (28, primary pain, 10 years) and partner (31) *

<table>
<thead>
<tr>
<th></th>
<th>NRS</th>
<th>FSFI</th>
<th>MPI_S</th>
<th>MPI_P</th>
<th>MPI_F</th>
<th>PHQ9</th>
<th>GAD7</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion</td>
<td>8</td>
<td>17</td>
<td>25</td>
<td>6</td>
<td>27</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>5(-15)</td>
<td>25.10(23.82)</td>
<td>23(-6.67)</td>
<td>5(-5)</td>
<td>26(-3.33)</td>
<td>7(14.81)</td>
<td>8(14.28)</td>
<td>9/10</td>
</tr>
<tr>
<td>Follow-up</td>
<td>2(-15)</td>
<td>27.10(5.88)</td>
<td>26(10)</td>
<td>5(0)</td>
<td>27(3.33)</td>
<td>5(-7.67)</td>
<td>1(-33.33)</td>
<td>9/10</td>
</tr>
<tr>
<td>Totalb</td>
<td>-30%</td>
<td>29.70%</td>
<td>3.33%</td>
<td>-5%</td>
<td>0%</td>
<td>7,41%</td>
<td>-19.04%</td>
<td></td>
</tr>
<tr>
<td>Partner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inclusion</td>
<td>20</td>
<td>4</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>30(33.33)</td>
<td>4(0)</td>
<td>27(6.67)</td>
<td></td>
<td></td>
<td></td>
<td>7/10</td>
<td></td>
</tr>
<tr>
<td>Follow-up</td>
<td>missing</td>
<td></td>
<td>6(10)</td>
<td></td>
<td>28(3.33)</td>
<td></td>
<td>7/10</td>
<td></td>
</tr>
<tr>
<td>Totalb</td>
<td>10%</td>
<td>10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note.  NRS = Numerical Rating Scale for pain, FSFI = Female Sexual Function Index, MPI S = MPI-SR solicitousness scale.  MPI P = MPI-SR punishing scale.  MPI F = MPI-SR facilitating scale.  TS = Treatment satisfaction.  Scores in brackets = % change since previous measure occasion in relation to full scale score.

*Compliance: group, 90% (did not attend session 3), partner sessions, 100%.

b in relation to full scale score.

Were detected in avoidance throughout. Catastrophizing displayed stability in the baseline phase, but instability (≤ 1.52 SD) in the treatment phase. There was a large degree of overlap in data from baseline and treatment (catastrophizing 93.75%, fear 75%, and avoidance 80.25%). It was concluded that woman 3’s catastrophizing, fear and avoidance deteriorated during the treatment phase.

PEM for catastrophizing (.19), fear (b, .25, p < .001) and avoidance (b, .19, p < .001) supported deterioration during treatment (Figure 5., b). All PEM scores indicated ineffective treatment. PEM4 for catastrophizing indicated instability throughout (Figure 5., c). PEM4 indicated that increases followed by decreases occurred catastrophizing and fear, while
avoidance remained stable. This suggests that avoidance may follow a different process tangent for woman 3.

**Couple 4.**

*Primary outcome measures.* Pain displayed a slight downward trend in the baseline and no trend in the treatment phase for woman 4. Directional change occurred at phase break. Function displayed an upward trend in the baseline, but no trend in the treatment phase. Directional change occurred at phase break. C-f-a displayed a slight downward trend in the baseline and no trend in the treatment phase (Figure 6., a). Directional change occurred in the third week of treatment. Pain and c-f-a displayed stability throughout. Function displayed stability in the baseline, but instability (> 1SD) in the treatment phase. Changes in medians over 1 SD were not detected in pain, function of c-f-a for woman 4. Some change in level between baseline and the treatment phase was observed function (-0.7 SD) and indicated deterioration. Level change was initially sustained but function was regained in later stages of treatment. There was a large degree of overlap data from baseline and treatment phases for all three variables (pain 68.75 %, function 87.5 %, c-f-a 87.5 % [Figure 6., a]). It was concluded that woman 4’s weekly measures of pain indicate small improvements during treatment. Weekly measures of function and c-f-a were seen to indicate negligible change or deterioration in treatment.

PEM pain (.63, small effect size) supported small improvements during treatment. PEM function (.18) and c-f-a (0.13) indicated deterioration during treatment (Figure 6., a). The trends seen in weekly measures reflected woman 4’s inclusion and post measures of pain and function (Table 5). Woman 4 reported a 50% total decrease in pain (15 % during treatment) and a 7.35% increase in function. Larger gains were reported in the FU period than during treatment for both variables.
Figure 6. Weekly measures of pain, function, and combined catastrophizing, fear, and avoidance; and weekly measures of catastrophizing, fear, and avoidance, for woman 4. Dashed lines indicated the median of the previous phase. Decimal scores denote percentage improvement compared to the median of previous phase (PEM). Fisher’s exact test is reported where $p \leq .05$. Black phase break denotes baseline to treatment. Red phase break denotes end of treatment.
Solicitous responses were unanimously reported to have increased during treatment, (woman 20%, man 30%) with small declines at FU. Punishing responses were unanimously reported to be very low or absent throughout, but man 4 reported fluctuations of ± 10%. Facilitative responses are reported heterogeneously. Woman 4 reported stability and very high facilitative responses throughout. Man 4 reported a gradual increase in facilitative responses from high at inclusion to very high at FU (total 20% change).

Table 5

<table>
<thead>
<tr>
<th></th>
<th>NRS</th>
<th>FSFI</th>
<th>MPI_S</th>
<th>MPI_P</th>
<th>MPI_F</th>
<th>PHQ9</th>
<th>GAD7</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inclusion</strong></td>
<td>15</td>
<td>23.40</td>
<td>27</td>
<td>4</td>
<td>36</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Post</strong></td>
<td>12(-15)</td>
<td>23.80(1,18)</td>
<td>33(20)</td>
<td>4(0)</td>
<td>36(0)</td>
<td>0(0)</td>
<td>0(0)</td>
<td>10/10</td>
</tr>
<tr>
<td><strong>Follow-up</strong></td>
<td>5(-35)</td>
<td>25.90(6,17)</td>
<td>30(-10)</td>
<td>4(0)</td>
<td>36(0)</td>
<td>0(0)</td>
<td>0(0)</td>
<td>9/10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>-50%</td>
<td>7.35%</td>
<td>10%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td><strong>Partner</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inclusion</strong></td>
<td>17</td>
<td>4</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Post</strong></td>
<td>26(30)</td>
<td>6(10)</td>
<td>29(6.67)</td>
<td></td>
<td></td>
<td></td>
<td>7/10</td>
<td></td>
</tr>
<tr>
<td><strong>Follow-up</strong></td>
<td>23(-10)</td>
<td>4(-10)</td>
<td>33(13.37)</td>
<td></td>
<td></td>
<td></td>
<td>7/10</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>20%</td>
<td>0%</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** NRS = Numerical Rating Scale for pain, FSFI = Female Sexual Function Index, MPI_S = MPI-SR solicitousness scale. MPI_P = MPI-SR punishing scale. MPI_F = MPI-SR facilitating scale. TS = Treatment satisfaction. Compliance = % sessions attended. Scores in brackets = % change since previous measure occasion in relation to full scale score.

Compliance: Group, 100%, partner sessions, 100%.

Secondary outcome measures. Woman 4 and her partner had complete treatment compliance. Woman 4 considered the treatment to have alleviated her pain moderately at both post treatment and FU. This is supported by her NRS scores (Table 5). Her partner considered her to have had small improvements on both occasions. Woman 4 considered
there to be a large improvement in the quality of her sex life on both occasions. This is not supported by her FSFI scores (Table 5). Man 4 considered there to be small improvements in his sex life at both post and FU. Woman 4 reported no symptoms of depression or anxiety throughout the whole study period.

**Process measures.** Catastrophizing displayed a steep downward trend in the baseline but a slight upward trend in the treatment phase for woman 4 (Figure 6., b). Directional change occurred at phase break. Fear displayed a slight downward trend in the baseline that was continued into the treatment phase, with a slightly steeper slope. Latency of change was unclear. Avoidance displayed no clear trend throughout. Catastrophizing and avoidance displayed instability (> 1SD) throughout. Fear displayed stability throughout. Change in median nearing 1 SD was observed in avoidance (0.93 SD) for woman 4, but not in catastrophizing or avoidance. Change in level nearing 1 SD between baseline and treatment was observed in avoidance (0.93 SD). There was a large degree of overlap in data from baseline and treatment for catastrophizing (100%) and avoidance (87.5%). There were few overlapping data points between phases for fear (25 % [Figure 6., b]). It was concluded that woman 4’s weekly measures of fear indicated improvement during treatment, but weekly measures for catastrophizing and avoidance indicated deterioration or no change.

PEM fear (.75, moderate effect size) supported improvements during treatment. PEM catastrophizing (0, \( p = .004 \)) and PEM (.13, \( p = .003 \)) supported significant deterioration during treatment (Figure 6., b). PEM4 (Figure 6., c) scores suggested that changes occurred in avoidance and catastrophizing simultaneously. PEM4 fear did not follow the same process tangent as catastrophizing and avoidance for woman 4.

**Couple 5.**

**Primary outcome measures.** Pain displayed no trend in the baseline, but slight upward trend, followed by a downward trend in the treatment phase for woman 5. Increases
occurred in the first three weeks of treatment, followed by slight and then steep decline. Large reductions were seen by the seventh week of treatment. Function displayed no trend in the baseline, but a slight upward trend in the treatment phase. Directional change occurred in the fourth week of treatment. C-f-a displayed a downward trend in the baseline which was continued into the treatment phase (Figure 7., a). In the fourth week of treatment c-f-a entered a steeper downward trend. Pain and c-f-a displayed stability throughout. Function displayed instability (> 1SD) in both phases. Function also displays missing data due to technical errors in connection with periods of instability. Changes in medians over 1 SD were detected in pain (-1.48 SD) and c-f-a (-1.42 SD) for woman 4. Large changes in median were not detected in function. Change level > 1 SD between baseline and the treatment phase were not observed for any variable, but changes in c-f-a were notable (-0.68 SD [Figure 7., a]). There was little overlapping data between baseline and treatment phases for all three variables (pain 25 %, function 21.43 %, c-f-a 12.5 % [Figure 7., a]). It was concluded that woman 5’s weekly measures of pain and c-f-a indicate large improvements during treatment. Weekly measures of function were seen to indicate small improvements in treatment.

PEM pain (.88, moderate effect size, p < .001) and c-f-a (1, large effect size, p = .002) support moderate to large improvements during treatment. PEM function (.93, large effect size, p = .006) support improvements during treatment (Figure 6., a). The effect size for function is large, but this is not reflected in the magnitude of change. The trends seen in weekly measures reflect woman 4’s inclusion and post measures of pain and function (Table 6). Woman 5 reported a 50% total drop in pain. Data for function was missing at post and FU. Reductions in pain were incurred during treatment and sustained at FU (see Table 6).

Solicitous responses were reported somewhat heterogeneously. Woman 5 reported a 16.67% increase at post treatment, but this was not sustained at FU. Man 5 reported stability. Punishing responses were uniformly reported to have decreased (total decreases, woman -
**Figure 7.** Weekly measures of pain, function, and combined catastrophizing, fear, and avoidance; and weekly measures of catastrophizing, fear, and avoidance, for woman 5. Dashed lines indicated the median of the previous phase. Decimal scores denote percentage improvement compared to the median of previous phase (PEM). Fisher’s exact test is reported where \( p \leq .05 \). Black phase break denotes baseline to treatment. Red phase break denotes end of treatment.
Facilitative responses were reported heterogeneously. Woman 5 reported stability (post treatment data is missing). Man 5 reported consistently higher levels of facilitative responses throughout and a 16.67% increase at post treatment that was partially sustained at FU.

**Secondary outcome measures.** Woman 5 and her partner had a good level of treatment compliance. Woman 5 considered the treatment to have alleviated her pain a lot at both post treatment and FU. This is supported by her NRS scores seen in Table 6. Her partner considered her to have had large improvements at post and small improvements at FU. Woman 5 considered there to be a large improvement in the quality of her sex life at

---

### Table 6

**Scale scores and percentage change for woman 5 (37, secondary pain, 17 years) and partner (52)**

<table>
<thead>
<tr>
<th></th>
<th>NRS</th>
<th>FSFI</th>
<th>MPI_S</th>
<th>MPI_P</th>
<th>MPI_F</th>
<th>PHQ9</th>
<th>GAD7</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion</td>
<td>12</td>
<td>23,70</td>
<td>16</td>
<td>9</td>
<td>20</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>2(-50)</td>
<td>missing</td>
<td>21(16.67)</td>
<td>9(0)</td>
<td>missing</td>
<td>3(-3.70)</td>
<td>5(14.29)</td>
<td>9/10</td>
</tr>
<tr>
<td>Follow-up</td>
<td>2(0)</td>
<td>missing</td>
<td>16(-16.67)</td>
<td>7(-10)</td>
<td>18</td>
<td>6(11.11)</td>
<td>9(19.05)</td>
<td>9/10</td>
</tr>
<tr>
<td>Total⁷</td>
<td>-50%</td>
<td>0%</td>
<td>-10%</td>
<td>-6.67%</td>
<td>7.41%</td>
<td>33.33%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Partner**

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion</td>
<td></td>
<td>18</td>
<td>16</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>17(-3.33)</td>
<td>10(-30)</td>
<td>32(16.67)</td>
<td></td>
<td>10/10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow-up</td>
<td>18(3.33)</td>
<td>9(-5)</td>
<td>29(-10)</td>
<td></td>
<td>5/10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total⁷</td>
<td>0%</td>
<td>-35%</td>
<td>6.67%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** NRS = Numerical Rating Scale for pain, FSFI = Female Sexual Function Index, MPI S = MPI-SR solicitousness scale. MPI P = MPI-SR punishing scale. MPI F = MPI-SR facilitating scale. TS = Treatment satisfaction. Scores in brackets = % change since previous measure occasion in relation to full scale score.

⁷Compliance: group, 70% (did not attend sessions 3, 6 and 10), partner sessions (100%).

²in relation to full scale score.
post treatment and moderate improvement at FU. Her partner considered there to be large improvements in his sex life at post treatments and no improvement at FU.

Woman 5 reported negligible symptoms of depression and anxiety at inclusion (see Table 6). At FU symptoms of both depression and anxiety were reported to have increased slightly. As anxiety appeared to follow an upward trend throughout it may be that symptoms of anxiety were influenced by participation in the study.

**Process measures.** Catastrophizing displayed a downward trend in the baseline and this was continued in the treatment phase, before reaching a stable horizontal tangent in the eighth week of treatment (week 14). Fear displayed a slight downward trend throughout. Avoidance displayed no trend in either phase. Catastrophizing displayed stability throughout. Fear displayed instability (> 1SD) in the baseline and showed signs of instability in the first four weeks of treatment. Stability was displayed in fear thereafter. Avoidance displays instability (> 1SD) in both phases, but this is confined to two episodes where levels of avoidance peaked for two weeks before returning to a stable plane. Changes in median > 1 SD were detected in catastrophizing (b, -1.67 SD), and fear (b, -1.81 SD [Figure 7., b]). Smaller changes were observed in the median of avoidance (-0.56 SD). Change in level > 1SD between phases was observed only in fear (-1.55 SD). There a little overlap in data from baseline and treatment for any variable (catastrophizing 12.5%, fear 12.5%, and avoidance 12.5%). Woman 5’s weekly measures of catastrophizing, fear and avoidance were seen to indicate improvement in the treatment phase.

PEM catastrophizing (1, large effect size, $p = .002$) fear (1, large effect size, $p = .01$) and avoidance (.88, moderate effect size, $p < .001$) supported significant improvement during treatment (Figure 7., b). PEM4 (Figure 7., c) scores indicated covariance between catastrophizing and fear, suggesting that avoidance may follow a different process tangent.
Discussion

Findings related to the primary aim

The primary aim of this thesis was to examine the feasibility of the proposed CBT+ group treatment for provoked vulvodynia. Feasibility was defined as treatment gains in pain intensity, sexual function, and partner response (increases in facilitative and decreases in solicitous and punishing responses). Within this aim was an intention to assess compliance, completion, perceived credibility, and change in depression and anxiety.

This study found some support for the hypothesis that the treatment would reduce women's pain. While there are discrepancies between weekly measures and post treatment measures of pain (e.g., woman 3), four of five women reported improvement of up to 60% from inclusion to post treatment, and FU. Visual analysis and PEM supported deteriorations for woman 3 in weekly measures, and woman 1, who also reported deterioration at post and FU. Effect sizes for the remaining women were small to moderate. This replicates earlier effect sizes for pain treatments (e.g., Kamper et al., 2015). Both woman 1 and 3 reported primary pain, suggesting the treatment may be less effective for primary vulvodynia.

According to Dworkin et al., (2008), results at FU indicate meaningful change. Improvements in pain are also in line with Corsini-Munt et al. (2014) who found that couples CBT for PVD reduced pain by ca. 50% when measured by NRS.

Results regarding sexual function are inconclusive and findings reflect those of ter Kuile and Weijenborg, (2006). Visual analysis of weekly measures of function identified variability in the treatment phase for women 2, 3, and 4, but stability for women 1 and 5. Furthermore, PEM indicated that the treatment had no effect on sexual function for woman 1, 3 or 4. While PEM indicated moderate effect sizes for woman 2, the magnitude of change was negligible. Improvements in function were seen in woman 5’s weekly measures (moderate effect size), but again, these were small in magnitude. Post treatment sexual
function data was only available for four women (not woman 5). Three women reported improvement at post (range 1.18-23.82%) but woman 1 reported deterioration. At FU, further small improvements were reported by women 3 and 4, but woman 2 had lost half of her gains, and woman 1 had deteriorated further. This suggests that gains in function were less stable than those in pain. None of the women attained sexual function approaching that seen in control groups, and women 1 and 2 consistently reported levels under that found indicative of vulvodynia (Masheb et al., 2004; Rosen et al., 2000). Thus, no clear conclusions can be drawn about treatment value for sexual function.

The hypothesis regarding frequency of solicitous, punishing and facilitative partner responses was not supported. Most participants reported increases in the frequency of solicitous responses during treatment. This suggests that the treatment may have iatrogenic effects but may also reflect maturation and testing effects. Most participants reported marginally increased frequency of facilitative responses. Floor effects were observed in punishing responses for all participants except couple 5, who reported decreases. Thus, the effects of treatment on pain, sexual function and partner response are ambiguous.

**Compliance, completion, and perceived credibility.** Ambiguous results were also found regarding compliance, completion, and perceived credibility. One of six couples who started treatment dropped out (16.67%), but levels of compliance, completion, and perceived credibility were good to excellent for all completers. However, there were discrepancies between women’s post treatment perceptions of pain relief and sex-life quality, and post treatment measurements. Most women perceived larger improvements than NRS and FSFI scores suggest. Male partners’ perceptions of improvements were better aligned with women’s scores. Further, post treatment NRS and FSFI scores appeared to be inflated in relation to weekly measures (e.g. woman 1). This may indicate social desirability bias or suggest that women do not necessarily value the quality of their sex lives based on the
functional domains of the FSFI. These seemingly contradictory results could be explained by the lack of other available treatments. Participants may have viewed the trial as their only hope for symptom relief. However, testing effects, or the inability of weekly measures to capture meaningful change could also explain these results. Thus, results were taken to indicate that the treatment is acceptable for participants.

**Depression and anxiety.** No clear support was found for change in these variables during treatment. Patterns of change were not replicated between participants. This opposes Masheb, Kerns, Lozano, Minkin, and Richman (2009) findings that individual CBT for PVD was successful at reducing symptoms of depression. The group format of this treatment may explain this. Additionally, Masheb et al., (2009) measured depression using the Beck Depression Inventory meaning results may vary due to instrumentation. Additionally, low symptoms at inclusion could mean that floor effects masked improvements in this study. In sum, evidence was not found of meaningful effects on depression or anxiety.

While results are mixed, they cumulatively suggest the treatment shows some promise. Change was observed before the eighth session for some participants (e.g., 2 and 5), and the second couples’ session was identified as potentially important in the catalysation of change (e.g., woman 5). Additionally, high levels of cohesion and completion indicate that the treatment is acceptable, and worth further development.

**Findings related to the secondary aim**

The secondary aim of this thesis was to examine the pattern of change in catastrophizing, fear, and avoidance. It was hoped that the process of return to function would be illuminated. It was hypothesized that the pattern of change would fit the FAM, with changes in catastrophizing preceding changes in fear, preceding changes in avoidance. As weekly measures of pain and sexual function were not indicative of significant improvements for most women it was not possible to draw reliable conclusions about return
to function. Notwithstanding, woman 5 did report improvement on both variables. Visual analysis of process variables and PEM4 scores for woman 5, however, provide no clear support for the pattern of the FAM. Neither do they support the pattern of change proposed for exposure treatments (avoidance, fear, catastrophizing [e.g., Craske et al., 2014]). Rather, avoidance was seen to co-vary primarily with function. Additionally, while instability was seen in fear during the first 14 weeks for woman 5, catastrophizing followed a smoother downward tangent. Moreover, woman 5’s reductions in pain at week 13 were not seen to co-vary with any other variable, although stability was reached at this point. The lack of replication, however, disallows conclusions about return to function.

As weekly measures indicate deterioration for many participants, the results did allow comparison with the pattern of the FAM. However, visual analysis, PEM and PEM4 led to ambiguous results. Although changes in catastrophizing were seen to precede changes in fear and avoidance for some women (e.g., 1 and 2) this was not seen in all. Additionally, changes in fear and avoidance reported by woman 1 were not sequential. Furthermore, woman 4 saw some simultaneous changes in avoidance and catastrophizing. However, a degree of covariation was seen in at least two c-f-a variables for most participants. It should be noted that in several cases covariation was between avoidance and another variable (e.g., woman 1, 2, and 4). This suggests that covariation was not exclusively a product of selecting the catastrophizing and fear items from the same subscale. Additionally, in most cases, one c-f-a variable was seen to follow a different processual tangent (e.g., fear; woman 4). It was however noted that covariation was seen between function and the combined c-f-a variable for most women (not woman 5). Moreover, covariation between function and c-f-a was seen to be greater than between either variable or pain. This could be indicative of the relationship between catastrophizing, fear, avoidance, and function/dysfunction that the FAM proposes.
It cannot however, be seen as supporting pattern fit. Thus, this thesis found no clear support for the pattern of the FAM.

There may be several possible explanations for the lack of support for the pattern of the FAM found in this thesis. The items used to measure catastrophizing, fear and avoidance were selected post hoc, and were not optimal for capturing each construct. These items also had small scales (4-7 items), meaning that they were not able to illustrate subtle changes. Additionally, as no “gold standard” for the analysis of process data on pattern fit was found, analysis methods were improvised. Thus the 4 week time scale employed in PEM4 was arbitrary. It was however chosen because (a) it was pragmatic and allowed for neat presentation and comparison of equal periods, and (b) it is a smaller measurement gap than regularly used in group studies. Thus, while PEM4 allowed for comparison of treatment periods, it was not a sensitive measure. However, this reduction in time between measurements is justified as, in research, gradual refinements should be preferred unless otherwise indicated. Furthermore, the use of digital measures with constant access meant that the length of time between measurements was not uniform. There is however, no consensus on the timeframe in which the sequence of the FAM takes place. Additionally, many of the earlier studies which have proposed sequentiality through mediation (e.g. Lane et al., 2018; ter Kuile, et al. 2015) have used intervals of 6+ weeks. Therefore, weekly measures and PEM4 were a refinement. Bi-weekly, daily, or more frequent measures may have provided clearer results. To summarize, although the data may be considered flawed, this study may indicate that there may be patterns of change other than that suggested by the FAM.

**Theoretical Implications**

The theoretical implications of this thesis lie in the lack of support found for the pattern of the FAM in this data set. While weekly measures do support covariance between elements of the FAM, no clear evidence was found to support pattern fit. Thus, a critical re-
evaluation of the FAM may be called for. Proposals such as alternate pathways (e.g., Pincus et al., 2010; Ward & Thorne, 2006) could help explain why differentiated tangents of development were found in c-f-a variables in this study. Lack of pattern fit may also indicate that other variables are integral in the pain cycle (if it is a cycle). As both Wideman et al. (2013) and Vlaeyen, Crombez and Linton (2016) propose, it may be necessary to further consider the importance of context. It may not be the micro-level captured by the FAM that has the greatest influence on the course of chronic pain. It may be that inter-relational factors on the meso-level play a larger maintenance role than previously thought. This could explain why couple’s treatments for PVD (e.g., Corsini-Munt et al., 2014) produce such promising results. Furthermore, findings of Lane et al. (2018), and Reme et al. (2012) suggest that systemic macro-level factors may play an integral role (e.g., structure of healthcare systems and communication). As such it may be motivated to widen the FAM so that meso and macro level factors become an understandable part of the process.

The results of this study may also suggest that a paradigm shift away from the FAM is needed to develop more effective pain treatments. Cumulative risk factor scores have been found to be a better predictor of future back pain than combined scores of catastrophizing, fear and avoidance (Wideman et al. 2012). Further, Wideman, et al. (2013) argue that development of a multidimensional framework with cumulative weighed risk factors may enable the development of innovative treatments. Such a frame-work may also forward understanding of pain and related disability from a truly biopsychosocial perspective. Wideman, et al. (2013) excuse their lack of model outline with space limitations. Nor was it within the scope of findings of this thesis to propose a multidimensional model.

**Clinical Implications**

Although outcomes other than for pain are somewhat disappointing, the treatment does show promise. Immediate effects were not expected, and visual analysis showed that
change occurred for some participants before the eighth treatment session (e.g., women 2 and 5). Change seen in this time span have previously been shown to indicate outcome (e.g. Lambert & Bergin, 2013, p. 204). These results suggest the treatment deserves development.

The disparity between women’s treatment satisfaction of reported improvements in sexual function may also have clinical implications. Results such as those reported by woman 1 appear surprising. However, they may indicate that penetrative sex is of lesser importance for a satisfying sex life that the treatment implies. This means that treatments may do better to focus on widening sexual repertoires. It must, however, be noted that woman 1’s male partner did not perceive improvements in his own sex life, and in general, men reported smaller improvements than women. Additionally, earlier qualitative research has shown that penetrative sex is normative and thus highly valued (e.g., Elmerstig, 2009). This supports pain reduction and successful penetrative intercourse as appropriate goals. However, researchers and clinicians have a responsibility to consider whether penetrative sex is a functional norm to continue supporting.

The lack of pattern fit with the FAM found in this data may also help to explain why psychoeducation has previously been found to be unappreciated by treatment participants (Rosen, 2018). Psychoeducation has been shown to have little effect in other pain treatments (e.g., Foster, Taylor, Eldridge, Ramsay & Griffiths, 2007). If the pattern of the FAM does not hold, it may be that participants gained no insight from psychoeducation as they did not recognize themselves in the model. As several earlier researchers have found evidence of alternate pathways (e.g., Flink et al, 2010; Ward & Thorne, 2006; Wideman et al., 2009), FAM for psychoeducation may benefit from variations presenting alternate pathways. Moreover, if the pattern of the FAM does not fit, it is likely that treatment interventions based on the model are suboptimal in focus and order. The results regarding pattern fit found in this thesis may therefore have wide reaching clinical implications.
Avenues for Future Research

The theoretical implications of this thesis motivate future research on return to function and pattern fit of the FAM. Ethical SCED studies of the development of chronic pain may be possible using currently pain free individuals awaiting surgery. Further SCED studies of ongoing pain, and return to function, with the primary aim of identifying the process of change are also recommended. By successively closing the measurement gap, it may be possible to find real-time evidence of pattern fit to the FAM. Alternatively, such studies may disconfirm pattern fit and further motivate a shift to a more complex model of chronic pain.

Only women with primary pain (1 and 3) showed increased pain in weekly measures. This suggests that there may be differences in the way in which treatment mechanisms effect women with primary pain. Further research is needed to discern whether this result was due to chance, or whether women with primary pain systematically differ from those with secondary pain in response to this treatment protocol. As such, the protocol may need further development to aide women with primary pain.

Further recommendations for research on the treatment of PVD include development of partner sessions. For instance, validation techniques could be incorporated as focus should lie on the reduction of solicitous responses. Further development of partner sessions may also form a compromise between the cost effectiveness of group CBT and the advantageous results seen in couple's therapy for PVD (e.g., Bergeron, Merwin, Dubé, & Rosen, 2018). Additionally, treatment elements that have previously been shown to have transdiagnostic value for mood disorders could be incorporated into the treatment. To increase the generalizability of the proposed PVD treatment it is also recommended that the protocol is tested on a non-WEIRD sample. Further, if other future research continues to question the
validity of the FAM, a new model for psychoeducation may be needed. There are as such, several avenues for further research.

**Strengths and Limitations of this Study**

This thesis has several weaknesses. Firstly, the baselines were not randomized. This disallows control for history or selection bias. The lack of randomization also prevented the use of more sensitive statistical analysis (e.g. randomization tests). Additionally, while improvement during baseline, (e.g., woman 1 and 4) may be a sign of regression to the mean, it could also be explained by suboptimal recruitment. Recruitment procedures were identical for the SCED and RCT. As such, SCED participants received information about treatment content during screening, contaminating the baseline. Thus, there are several threats to internal validity in the baselines.

Further threats to internal validity can be found in the use of an RCT style questionnaire battery for weekly measures. Although single items were selected for the purpose of this study, female participants filled in a weekly measure consisting of 41 items. This suggest that testing effects, fatigue and maturation are probable threats. Further, as the weekly questionnaire was taken directly from the monthly RCT questionnaire, all items were answered based on “experience over past 4 weeks”. Results should thus, be composites of the past month, rather than week. Further, there is missing data, meaning that comparisons of FSFI and MPI were not possible for all participants (man 3 and woman 5). As both were primary outcome measures, this reduced the number of results on which conclusions about feasibility were based. Missing data was, however, due to problems with digital questionnaires, suggesting that fuller datasets may be attainable with replication. Thus, these threats to internal validity could be rectified with simple modifications and replication.

Further threats to internal validity may also be found in the use of self-report measures. Self-report is problematic as results are, by nature, subjective. Therefore,
proponents of single case research advocate behavioural process measures, as they are auditable, quantifiable, and comparable between participants (e.g., Morley, 2017, p. 46-47). However, this study aimed to measure process concepts with covert, as well as overt characteristics. This means that observational measures may not have captured the breadth of each concept. Additionally, the items chosen to represent fear and avoidance captured overt behaviours, meaning that an external observer could have audited results. The items were however fitted to the constructs in retrospect. They were therefore neither validated, or idiosyncratic measures tailored to each participant. As such, the items may be misconstrued.

For example; the avoidance item was the outermost item of the PBQ and covered only penetrative intercourse. The item was considered to be least contaminated by the treatment protocol. It may however have been insensitive to change, and floor effects were seen for several women (e.g., 5). Moreover, by only measuring avoidance of penetrative intercourse, it was not possible to gauge whether the treatment successfully reduced other avoidance or broadened women’s sexual repertoires. It is thus possible that process measures were not successful in capturing the intended concepts.

A further problem with the process measures is that items were taken from different scales, meaning scores could not be compared in raw form. Z-scores were used to allow comparisons between variables and create the combined variable, but visual impact of floor and ceiling effects was lost. A simple range conversion however, did confirm that graphs using z-scores did not differ considerably. Additionally, to aide visual understanding of floor and ceiling effects, graphs were made with a smaller negative than positive span on the y-axis. Thus, steps were taken to minimise the risk of misinterpretation due to z-scores.

A final threat to internal validity is that video recordings of treatment sessions had not been reviewed. It was therefore not possible to assess whether the treating psychologists had
faithfully followed the treatment protocol. In summary, there are several threats to internal validity meaning the results of this thesis should be viewed with caution.

Regarding external validity, the sample in this study was WEIRD (Henrich, Heine, & Norenzayan, 2010). There are therefore issues with generalizability. Whilst far from unique for psychology research (e.g., Rad, Martingano, & Ginges, 2018), it means that results in this thesis can only be applied to other WEIRD contexts. Further, considering the sensitive cultural nature of sexuality (World Health Organisation, 2006, p. 7), results may only be generalizable in Scandinavia. Additionally, women were ethnic majority, and had low frequency of comorbid pain and mood disorders. This suggests that the sample did not adequately represent risk groups for vulvovaginal pain (e.g., Harlow et al, 2014; Khandker et al., 2011). Together with the goal of penetrative sex, this can be seen as research fuelling Eurocentric norms, which Rad et al. (2018) caution against. The sample is, however, representative of those included in earlier studies of vulvovaginal pain, suggesting that results may be used for benchmarking.

Other limitations are related to the use of PEM. PEM provides scores which can be equated to effect size, which was important for the treatment hypotheses and comparison with effect sizes reported in earlier studies. However, PEM is not capable of distinguishing stability, trend, or magnitude, and high PEM scores do not necessarily entail meaningful change. This study used questions on comparatively small scales with large intervals (1-4 or 0-6). Thus, the semantic difference between scores may be considered larger than when large scales with smaller intervals are used (e.g., avoidance was answered; “No attempts made”, “Have attempted but not succeeded”, “Have been successful on some attempts”, “Have been successful on all attempts”). This may increase the validity of PEM as a measure of effect size, as “small” increases in scores should entail qualitatively different behaviours or experiences. Moreover, there is no universal agreement on proper calculation of probability
values for PEM (e.g. Morley, 2017, p 134-135). In summary, there are problems with the analyses used in this study, but better fitting options were not found.

This study has several strengths. Firstly, it was believed to be the first SCED examining a treatment for vulvovaginal pain. As such the results give unique insight into the process of change during treatment. Additionally, it is believed to be the only SCED which has aimed to examine pattern fit and return to function with reference to the FAM. Shadish et al. (2002) write that studies of pattern fit using time series data may be preferred to group studies, as they can simply negate alternative hypotheses. Additionally, Shadish et al., (2002, p. 484-485) suggest that the use of quasi-experimental designs for such purposes can be defended, as such studies provide invaluable information. This suggest that the design implemented in this study is both suitable and parsimonious.

An inherent strength of SCED studies is that possible placebo effects can be seen in the baseline. Some women in this study experienced improvement during the baseline. Such effects are likely to be invisible or assumed to be treatment effects in group studies. The baselines in this study also exceeded What Works Clearing House’s (2017) 5 point recommendation. Additionally, the use of multiple analyses helps to strengthen the validity of the results. It has been suggested that visual analysis reduces the risk of Type I errors, as large effects are necessary for the detection of change (Parsonson & Baer, 1978). However, there is no accepted method for judging the risk of making Type II errors and contrary to Parsonson and Baer (1978), there is evidence that analysts often find non-existent effects through visual analysis alone (Kazdin 2011, p.300). Further strengths are that the author was blind to the participants, and the independent analyst was blind to the hypothesis, intervention and variables, which reduced the risk of analysis bias. Thus, this thesis has several strengths.
Conclusion

In conclusion, this thesis found that the CBT+ treatment protocol for provoked vulvodynia shows promise, but pattern fit with the FAM was not supported. Outcomes for pain were encouraging, with post and FU measures of pain showing meaningful change. Effect sizes for weekly pain measures were small to moderate, but this was expected considering earlier pain treatment research. Further refinements are recommended, with specific focus on increased sexual function and the reduction of solicitous partner responses. Regarding pattern fit to the FAM, further research, and emphasis on return to function is needed. Moreover, the results of this thesis suggest that researchers and clinicians may need to re-evaluate the theoretical grounds on which they currently base pain treatments. By exploring other research avenues, it may be possible to develop superior models and treatments.

In sum, there is work to do.
References


Allen, R., & Williams, A. (2001). Review: Behavioural treatment has a small to moderate beneficial effect on chronic low back pain. *Evidence Based Mental Health, 4*(2), 55-55. doi:10.1136/ebmh.4.2.55


GROUP CBT+ FOR PROVOKED VULVODYNIA


Hasenbring, M. I., & Verbunt, J. A. (2010). Fear-avoidance and Endurance-related Responses to Pain: New Models of Behavior and Their Consequences for Clinical


item response theory. *PLoS One, 12*(8), e0182162.

doi:10.1371/journal.pone.0182162


Appendix

Other Measures Taken but not Included in Analysis

The following measures were also taken at baseline, completion, and FU but not included in this study: the Approach and Avoidance Sexual Goals questionnaire (Cooper, Shapiro, Powers, 1998; Rosen, Dewitte, Merwin, & Bergeron, 2017); CHAMP Sexual Pain Coping Scale (Flink, Thomtén, Engman, Hedström, & Linton, 2015); EuroQol-5 dimensions (not at inclusion) (Herdman, et al., 2011); Female Sexual Distress Scale (Derogatis, Rosen, Leiblum, Burnett, & Heiman, 2002); Genital Pain Rating Questionnaire (Brauer, ter Kuile, Laan, & Trimbos, 2008); Global Measure of Sexual Satisfaction Scale (Lawrence & Byers, 1995); International Index of Erectile Dysfunction (Rosen, Riley, Wagner, Osterloh, Kirkpatrick,& Mishra, 1997); Male Sexual Distress Scale (Santos-Iglesias, Mohamed, Danko, & Walker, 2018); Maudsley Marital Questionnaire (Busby, Christensen, Crane, & Larson,1995; Crowe, 1978); Painful Intercourse Self-Efficacy Scale (Desrochers, et al., 2009); Sexual Physical Abuse Questionnaire (Kooiman, Ouwehand, & ter Kuile, 2002).