Universiteit Utrecht

The mismatch study: exploring the discrepancies between clinician's diagnosis, structured interview outcomes and patient self-report in PTSD

Kiki Meijer (3987728)

Master thesis Clinical and Health psychology

Faculty social and behavioral sciences

University of Utrecht

Supervisors: Dr. D. Cath and Dr. M. Hagenaars

August 2016

In collaboration with Altrecht Academic Anxiety center, Utrecht, the Netherlands

Abstract (EN)

A valid and reliable PTSD diagnosis is requisite to give PTSD patients the treatment they need. Therefore, obtaining the right diagnosis is crucial. Multiple studies demonstrate that discrepancies exist between the several diagnostic instruments and their outcomes. These discrepancies, or 'mismatches', might be partially explained by patient as well as clinicians' factors. The objective of the current study was to explore the amount of mismatch between the PSS-SR (self-report), SCID-I (standardized measure) and the clinician's diagnosis (clinical interview), as well as the possible explanations for this mismatch. This was done by performing a two-stage project.

Study 1. The first stage was the pilot-study. Data were obtained from the Jelgersma Clinic (Oegstgeest, The Netherlands). In a sample of borderline patients (N=28) only a fair agreement (Cohen's Kappa .267) was found between the clinician's diagnosis and the PSS-SR, with the PSS-SR diagnosing PTSD more often than the clinicians' judgment. Furthermore, the pilot study indicated that someone suffering from additional anxiety symptoms is more likely to be diagnosed with PTSD according to the PSS-SR.

Study 2. The objective of Study 2 was to replicate and extend the findings from study 1 in a large scale group, with additional structured interview data on PTSS diagnoses, by investigating the amount of mismatch between the PSS-SR, SCID-I and the clinician's diagnosis. As a second objective, explanations were sought at the level of patient and clinician characteristics for potential mismatch. The participants for Study 2 were all newly presented patients at the Altrecht Academic Anxiety centre (Utrecht, The Netherlands) who received a primary PTSD diagnosis according to the SCID-I and on whom data were available on clinicians' diagnosis at intake and on PSS-SR (N=184). Between the SCID-I and the PSS-SR a mismatch of 8% was found and between the clinicians' diagnosis and SCID-I a mismatch of 31% was found. No agreement (Cohen's Kappa .056) was found between the clinicians' diagnosis and the PSS-SR, with the PSS-SR diagnosing PTSD more often than the clinician. Study 2 did not find an indication that someone suffering from anxiety symptoms is more prone to receive a PTSD diagnosis. However, results indicated that suffering from depressive symptoms might enhance receiving a PTSD diagnosis according to the PSS-SR and the clinicians' opinion. Educational level of the clinician did not have any influence on the degree of mismatch. However, clinicians with 0-1 years and >20 years of experience showed the highest degree of mismatch (in comparison to the SCID-I), with respectively 50% and 42% mismatch. Overall, these findings are worrying with respect to the accuracy of diagnoses that clinicians make. The current project has been a first step to address this important issue.

Keywords: *mismatch, discrepancy, match, agreement, PTSD, clinical judgement, PSS-SR, SCID-I, diagnostic instruments, posttraumatic stress disorder, structured interview*

Samenvatting (NE)

Een betrouwbare en valide PTSS diagnose is nodig om iemand met PTSS te kunnen behandelen. Zodoende is het cruciaal dat iemand de juiste diagnose krijgt. Echter laten meerdere onderzoeken zien dat er niet altijd een juiste diagnose wordt gesteld. Er blijkt een discrepantie te zijn tussen de uitkomsten van de verschillende diagnostische methoden. Deze discrepantie, ofwel mismatch, kan mogelijk gedeeltelijk verklaard worden aan de hand van patiënt- en clinicus factoren. Het huidige doel betrof het onderzoeken van de mate van mismatch tussen de diagnoses van de verschillende diagnostische instrumenten: de klinische diagnose (klinisch interview), SCID-I (gestandaardiseerd interview) en de PSS-SR (zelfrapportage instrument). Daarnaast werden mogelijke verklaringen voor deze mismatch onderzocht. Huidig onderzoek bestaat uit twee fasen

Studie 1. waarbij de eerste fase een pilot-onderzoek betrof. De data is verzameld bij de Jelgersma kliniek (Oegstgeest, Nederland). In een steekproef van borderline patiënten (N=184) werd een *fair* overeenkomst (Cohen's Kappa .267) gevonden tussen de klinische diagnose en de PSS-SR, waarbij de PSS-SR PTSS vaker diagnosticeerde dan de clinicus. Daarnaast werden er aanwijzingen gevonden dat iemand die lijdt aan angstsymptomen een grotere kans heeft om met PTSS gediagnosticeerd te worden door de PSS-SR.

Studie 2. Het doel van de tweede fase was het repliceren en uitbreiden van studie 1 in een grotere sample, met een meer uitgebreide dataset. De mate van mismatch werd onderzocht tussen de PSS-SR, SCID-I en klinische diagnose. Daarnaast werden verklaringen gezocht voor de mismatches op het niveau van patiënt- en clinicus factoren. De steekproef van Studie 2 betrof de nieuw-aangemelde patiënten bij het Altrecht Academisch Angstcentrum (Utrecht, Nederland) wie als hoofddiagnose PTSS hadden gekregen op de SCID-I, er een klinische diagnose beschikbaar was uit de intake en van wie uitkomsten op de PSS-SR bekend waren (N=184). Er werd een mismatch van 8% gevonden tussen de klinische diagnose en SCID-I en een mismatch van 31% tussen de PSS-SR en de SCID-I. Er werd een mismatch van 32% gevonden tussen de klinische diagnose en de PSS-SR, wat inhoudt dat er geen overeenkomst is (Cohen's Kappa .056). De PSS-SR diagnosticeerde vaker PTSS dan de clinicus. Studie 2 vond geen aanwijzingen dat aanwezige angstsymptomen mogelijk leiden tot het sneller labelen van een PTSS diagnose. Daarentegen werden er wel aanwijzingen gevonden dat aanwezige depressieve symptomen mogelijk faciliterend zijn voor het verkrijgen van een PTSS diagnose door zowel de clinicus als de PSS-SR. Wat betreft clinicus factoren blijkt dat de mate van mismatch niet verschilt tussen de opleidingsachtergronden. Wel werd een verschil in mismatch gevonden wanneer gekeken werd naar jaren van klinische ervaring: clinici met 0-1 jaar en >20 jaar ervaring hebben de hoogste mate van mismatch – ten opzichte van de SCID-I – met respectievelijk 50% en 42% mismatch. De uitkomsten zijn zorgelijk, kijkend naar de accuraatheid van diagnoses gesteld door de clinicus. Huidig onderzoek is een goede stap in de richting in het aankaarten van dit probleem.

Preface

We worked closely together during the entire process and we exchanged our thoughts and ideas about the subject, the (sub)questions and the effectuation. We decided to explore the main subject – the mismatch between the different diagnostic methods – together and then divide the two areas of explanatory factors. Lotte decided to focus on the clinician level, and I decided to focus on the patient level. The parts Lotte wrote will be marked in green and the parts I wrote will be marked in yellow. The non-marked parts were mainly written together. Despite the division, we critically reviewed each others pieces regarding the explanatory factors and discussed everything through.

I would like to thank Muriel Hagenaars and Danielle Cath for guiding us through this process and for their critical view on the process and especially on the statistical analyses. We spent a lot of time discussing about the findings and what they could mean. I would like to thank the employees of Altrecht Academic Anxiety center as well, for collecting the data, preparing the dataset and for helping us out with everything and always responding quickly to our questions.
Most of all, I would like to thank Lotte – my partner in crime – for working together in this process and motivating each other. It was a great privilege to work this closely together and have the opportunity to discuss everything thoroughly, criticizing on each others work and helping each other out when one of us could not solve the puzzle.

After almost a year of hard work, I can honestly say I am satisfied with the result and with the outcomes of the study, I hope we shed some light on this matter and inspire other researchers to further investigate this issue. I have learned a lot: working together intensively, solving a puzzle, delivering a scientifically worthy thesis, upgrading my academic English writing skills, enduring setbacks and my patience have been put to the test. My career as a psychologists is about to start and I will keep the results of this thesis in mind when I start working in the field. I hope you will enjoy reading this thesis.

Index

Introduction

With more than two-third of the general population experiencing a significant traumatic event at some point in their lives, the occurrence of traumatic experiences is relatively common in humans (Galea, Nandi & Vlahov, 2005). Of all people who have experienced such an event, about 8% (Vermetten, 2014) will eventually develop a post-traumatic stress-disorder. Post-traumatic stress disorder (PTSD) was established as a psychiatric diagnosis in the DSM-III (American Psychiatric Association, 1980) and defined as a characteristic symptom pattern following exposure to a traumatic event. In the current DSM (5; American Psychiatric Association, 2013) this symptom pattern includes a triad of persistently re-experiencing the trauma (intrusions), negative alterations in cognitions, mood, arousal and reactivity and effortful avoidance of distressing trauma-related stimuli. These symptoms are associated with considerable suffering and functional impairment (Stein, McQuaid, Pesrelli, Lenox & McCahill, 2000). Persistent PTSD carries a high risk for chronicity (Kessler, Sonnega, Bromet, Hughes & Nelson, 1995). Furthermore, the economic burden associated with PTSD is high and specifically related to health care utilization (Stein et al., 2000; Wohlfarth, van den Brink, Winkel & ter Smitten, 2003). Because of the impairments in the person's quality of life and the high costs for the community and government, it is necessary that the right diagnosis is obtained, as it is to therapy within the framework of guideline development and the definition of guidelines in specific disorders (Andreas, Theisen, Mestel, Koch & Schulz, 2009).

Diagnostic methods

In psychiatry there are several ways to obtain a diagnosis (Van Balkom et al., 2013). The first way is the clinical interview. In mental health settings, the initial interview done by the clinician, using an unstructured, open-ended approach, remains the primary assessment tool for diagnosing mental disorders based on the DSM (Jones, 2010). In other words: the intake. Secondly, there are standardized diagnostic interviews (SDI's). SDI's were developed to operationalize diagnostic criteria and to increase the reliability and validity of diagnoses, by reducing: (a) information variance, i.e. basing diagnoses on different and relatively randomly provided information; (b) criterion variance, i.e. different clinicians defining disorders differently; (c) interpretation variance, i.e. different clinicians interpreting the same information differently (Rettew, Lynch, Achenbach, Dumenci & Ivanova, 2009). With a SDI, the diagnosis is obtained by having interviewers ask the same questions in the same order and then process the answers through standardized algorithms (Rettew et al., 2009). The third method is the use of self-report questionnaires (Van Balkom et al., 2013). Self-report questionnaires ask questions about symptoms and behaviors and their severity associated with mental disorders in a systematic and reproducible way in order to obtain insight into a patient's illness (Gregory, 2007). They are useful in giving direction to a DSM diagnosis, establishing symptom severity and they contribute to a complete understanding of a patient's diagnosis (Meyer et al., 2001). This method is

less time-consuming and therefore less expensive than the use of SDI's (Stice, Telch & Rizvi, 2000). However, they cannot replace structured interviews used to establish qualitative diagnoses but they are additive and extremely useful for *quantifying* illness severity. In Dutch clinical practice the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) (First, Spitzer, Gibbon & Williams, 1997) is the most commonly used SDI and is often seen as the 'golden standard' of diagnostic methods (Rettew et al., 2009). The PTSD Symptom Scale - Self-Report Version (PSS-SR) is an often used self-report questionnaire to assess PTSD symptom severity (Foa, Riggs, Dancu & Rothbaum, 1993; Engelhard, Arntz & van den Hout, 2007).

Discrepancies between the diagnostic methods

Despite the importance of diagnosing the right disorder, multiple studies demonstrate that discrepancies exist between the several diagnostic methods and their outcomes with respect to the final diagnosis given to someone (Rettew et al., 2009; Engelhard et al., 2007; Steiner, Tebes, Sledge & Walker, 1995; Kovess, Sylla, Fournier & Flavigny, 1992). This discrepancy or disagreement of a diagnosis based on different methods is called a 'mismatch'. In clinical practice however, the most desirable outcome is that the three available methods all agree and converge to one diagnosis and thus initiating a 'match'. The less favourable outcome is when the different methods lead to deviations with respect to the final result (mismatching). There are two ways in which a match can be obtained: when a disorder is correctly identified and when it is correctly rejected, respectively as a true positive and a true negative diagnosis. It is also possible to incorrectly identify a disorder (false positive) or incorrectly reject a disorder (false negative). The ideal outcome is when a method is capable of only diagnosing true positives and true negatives and not diagnosing any false positives or false negatives.

Several studies have looked at the different diagnostic methods and their matching abilities. Cohen's Kappa is used in these studies as the measure of agreement between raters or rating methods. The most used guidelines from the literature are those by Landis & Koch (1977) who introduced the following guidelines: a Cohen's kappa of 0 is seen as no agreement, 0-0.20 as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial, and 0.81-1 as almost perfect agreement.

The current knowledge concerning the agreement between clinicians, semi-structured interviews and self-reports

SCID-I vs. clinical diagnosis. In a patient sample of 100 persons with various diagnoses, Steiner and colleagues (1995) have detected an overall weighted Cohen's Kappa indicating fair agreement (0.25) between the SCID-I and the clinical diagnoses. They calculated the Cohen's Kappa for eight different diagnoses, ranging from 0 for panic disorder to 0.55 for schizophrenia. Unfortunately, they did not take PTSD into account. Rettew and colleagues (2009) did a meta-analysis

which included 38 articles in 15.967 subjects concerning the agreement between the SCID-I and a whole DSM range of clinical diagnoses established by clinicians. They found a Cohen's Kappa of 0.27 for the overall agreement which is considered fair. Of all subjects n=888 were included to obtain a Cohen's Kappa specifically for PTSD. A mean Cohen's Kappa of 0.54 was found, which is considered a moderate agreement, with the SCID-I detecting more cases of PTSD than the clinician. The mean Cohen's Kappa found across all anxiety disorders was 0.29 (fair).

Self-report (specifically the PSS-SR) vs. SCID-I. Engelhard and colleagues (2007) found that using the PSS-SR with a cutoff of 15, 86% of the individuals with PTSD, 5% of the healthy controls and 43% of the individuals with other anxiety disorders endorsed sufficient symptoms to meet a PTSD diagnosis. This was estimated with a sample of 65 individuals with PTSD, 40 individuals with other anxiety disorders and 40 controls. They conclude that the PSS-SR is a good instrument to distinguish healthy controls from PTSD patients. However, there were many false positives detected using the PSS-SR; 50% of the participants who had an anxiety disorder diagnosis (other than PTSD) met a PTSD diagnosis following the PSS-SR, whereas PTSD was clinically not present. Therefore the PSS-SR may not be a suitable diagnostic instrument in distinguishing PTSD from other anxiety disorders, possibly because of the presence of anxiety-related items on the PSS-SR. People suffering from another anxiety disorder theoretically score high on these anxiety sensitive items thereby increasing the total score on the PSS-SR which might lead to a total score that exceeds the cutoff score. With regard to the mismatch between PTSD diagnosis obtained from the SCID-I and the PSS-SR in this study, an overall agreement of 75% was achieved, with the PSS-SR diagnosing PTSD more often than the SCID-I. Unfortunately, Cohen's Kappa were not calculated in this study. Another study, Wohlfart and colleagues (2003) looked at the agreement between the PSS-SR and a standardized diagnostic instrument for establishing PTSD diagnosis, i.e.. the Composite International Diagnostic Interview (CIDI) They found an agreement of 85% which corresponds to a moderate Cohen's Kappa of 0.52. The sensitivity was 90% and the specificity 84%.

Clinical diagnosis vs. PSS-SR. Unfortunately, to date there is no literature available that looks into the agreement between the clinician's diagnosis and the diagnosis obtained from the PSS-SR.

Explanations for mismatch

At the *patient's* level, the study of Engelhard and colleagues (2007) showed that the PSS-SR lacks specificity (many false positives) in distinguishing PTSD from other anxiety disorders. Several explanations are given. First, the aversive events mentioned by many patients without PTSD are not typically classified as traumatic, at least not by the DSM-IV, but may bring forth symptoms that seem to be characteristic of PTSD. PTSD symptoms can arise from relatively ordinary stressors, such as problems in relationships, financial problems and troubles at work. Second, there is symptom overlap between other anxiety disorders and PTSD. For example, many hyperarousal symptoms of PTSD are

very similar to physiological anxiety-related symptoms such as excessive perspiration or increased heartbeat, symptoms that are frequently found in anxiety disorders. Further, PTSD-related symptoms of numbing are comparable to freezing symptoms in panic disorder. Third, several symptoms of avoidance behavior, numbing and hyperarousal are also present in depressive disorder. A depressed mood often occurs in anxiety disorders and PTSD. Which means that the high rates of false positive symptoms of PTSD in anxiety disordered patients could be explained by symptom overlap with depression. Meltzer and colleagues (2012) have found another explanation: when a patient with PTSD scores high on a depression scale, physicians are more likely to recognize these depressive symptoms, to ignore the underlying PTSD and to mislabel PTSD patients as having a depression as their primary diagnosis. These results indicate that comorbidity might be involved in the risk of a mismatch, especially when the comorbid symptoms concern a similar domain as the PTSD symptoms.

When seeking explanations at the *clinician* level, the first possible explanation is that clinicians do not always ask about important symptoms, do not weigh symptoms equally and do not stick to the diagnostic criteria during their interview of the patient (Garb, 2005; Kim & Ahn, 2002). For example, if a history of trauma is presented, clinicians may diagnose PTSD without probing all criteria (Rettew et al., 2009). Structured interviews are used to ensure that diagnoses are based on diagnostic criteria. Therefore, it is likely that the interrater reliability for clinical interviews is lower than diagnoses obtained with structured interviews (Miller, Dasher, Collins, Griffiths & Brown, 2001). Secondly, clinicians often do not receive sufficient feedback about the accuracy of their diagnosis and consequently fail to learn from experience (Sapyta, Riemer & Bickman, 2005; Dawes, 1989). The third explanation relates to cognitive heuristics and biases (Garb, 2005). Clinicians generally rely on limited numbers of heuristic principles which erroneously reduces the complex tasks of assessing probabilities and predicting values to simpler judgmental operations (Tversky & Kahneman, 1974). This results in clinicians making errors in the diagnostic process (Garb, 2005). Clinical judgment shows to be subjective and is based on a clinician's own beliefs and experiences.

One would believe that clinician's experience improves diagnostic accuracy. However, the literature on this subject is contradictory. Elstein and Schwarz (2002) propose that experts are faster and more efficient in composing a diagnosis than trained interviewers using SDI. On the other hand, experts use other reasoning processes than trained interviewers and novices. Experienced clinicians use a hypothetical-deductive strategy *only* with difficult cases and their clinical reasoning entails mostly pattern recognition or automatic retrieval. In contrast novices use a hypothetical-deductive strategy to compose a diagnosis most of the time. One might suggest that the more knowledge a clinician has, the more they are likely to use pattern recognition and automatic retrieval. The errors in reasoning (pattern recognition and automatic retrieval) include failure to generate the right hypothesis; misperceiving or misreading the evidence and misinterpreting the evidence. Moreover, Dawes (1989) argues that degree of experience in years fails to predict judgmental accuracy. As Dawes (1989)

defines: "I can tell on the basis of my experience with people of a particular type (e.g., child abusers) that this person is of that type (e.g., a child abuser)" are simply invalid."(p.457). As Dawes (1989) states, professional clinicians may make somewhat better judgments than non professionals, but these differences are not explained by the years of experience. They are mostly explained by differences in intelligence, or in terms of how they employ specific interview techniques. Garb (1989) argues in his review that experienced clinicians are not more accurate than less experienced clinicians. He gave two causes for this finding: clinicians do not often receive feedback, which makes it difficult to learn from experience; and errors in the cognitive processes, like wrong hypothesis testing and biases. Contrastingly, Spengler and colleagues (2009) have shown in their meta-analysis that years of experience do make the decision making capacity more accurate, with experience showing a small effect on judgment accuracy. In the existing literature there is still no clear evidence about the effect of experience on judgment accuracy (Garb & Boyle, 2003).

Besides clinical experience in years, it is interesting to look at the clinician's profession as a proxy of level of education. There has been explained in the previous paragraph that the better judgment of an experienced clinician might be a result of their knowledge on what techniques to use (Dawes, 1989). Assuming that clinicians with a higher level and more years of education have more mastery of content and have more diagnostic skills. Therefore, it seems of interest to investigate the clinician's background of training, hypothesizing that clinical psychologists and psychotherapists have more mastery of content and have more diagnostic skills than social workers or psychiatric nurses.

The current study

As mentioned, discrepancies between the different diagnostic methods and their outcomes exist for PTSD (Rettew et al., 2009; Andreas et al., 2008; Engelhard et al., 2007; Wohlfart et al., 2003; Foa et al., 1993). Due to the overall poor agreement, the divergent findings, the lack of extensive knowledge, the dated literature and the clinical implications of PTSD, it seems necessary to pay attention to the discrepancies of PTSD diagnoses and most importantly, to the possible explanations for these mismatches. The aim of the current study is to explore the mismatch between self-reported diagnoses, structured interviews and clinicians' diagnosis, in patients diagnosed with PTSD. A two stage project is performed. The first stage is a pilot study in which clinicians' diagnosis and self-reports of PTSD symptoms were available. The aim of the pilot study is to (1-1) explore the mismatch between the PSS-SR and the clinician's' diagnosis (yes/no) and (1-2) to explore whether severity of current anxiety symptoms serve as a possible explanation for mismatch. As there is no literature available concerning the agreement between the clinicians' diagnosis and the PSS-SR (PTSD self-report), no hypotheses can be formulated on the direction of the results. Regarding the severity of concurrent anxiety symptoms, it is expected that the presence of these symptoms will lower the agreement between clinician's diagnosis and self-report because increased scores on the anxiety -related items on the PSS-

10

SR might erroneously lead to false-positive self-reported PTSD diagnosis compared with clinicians' judgments (Engelhard et al., 2007).

The results of the first study will inform the questions for the second larger-scaled study. And extend the outcomes from the first study, in which data of PTSD patients regarding clinician's diagnoses, self-reports as well as structured interviews are available at the same timepoints before entering treatment. The first objective of study 2 (2-1) is to study the amount of agreement between the SCID-I (structured interview) and the PSS-SR (self-report); between the SCID-I and the clinician's diagnosis; and between the clinician's diagnosis and the PSS-SR. As a second objective, explanations for the potential mismatches are sought at the level of patient characteristics (anxiety symptoms, depressive symptoms) (2-2a) and clinician characteristics (years of experience, education level) (2-2b).

Hypotheses: (2-1) The agreement between the SCID-I and the clinician's diagnosis is expected to be poor to moderate (Steiner et al., 1995; Rettew et al., 2009; Andreas et al., 2008). Based on the literature, the agreement in percentages of patients' diagnostic match between the SCID-I and the PSS-SR (75%-86%) is expected to be moderate as well. As mentioned, there is no literature available concerning the agreement between the clinician's diagnosis and the PSS-SR. However, in line with the literature, we expect the match between PSS-SR cutoff derived diagnoses and SCID-I diagnoses to be higher than between PSS-SR cutoff derived diagnoses and clinician's diagnoses. Moreover, it is expected that the amount of mismatch will be similar to the amount of mismatch found in the pilot study.

Regarding the explanations for mismatch (between SCID-I and PSS-SR; between clinician's diagnosis and PSS-SR and between SCID-I and clinician's diagnosis) at the level of patients' characteristics (2-2a), it is expected that the higher the comorbid anxiety and/or depressive symptom scores, the lower the agreement between clinician's or SCID-I diagnoses and PSS-SR scores, because anxiety and depressive symptoms will inflate PTSD diagnoses on the PSS-SR (Engelhard et al., 2007; Meltzer et al., 2012).

Regarding explanations for mismatch at the level of clinician' characteristics (2-2b), we expect an effect of clinician's experience on diagnostic accuracy, although the specific direction of the effect is not clear given the contradictory findings in the literature. Finally, based on the literature, clinicians with higher educational levels are expected to be more accurate in their diagnoses. The study may provide new insights regarding the causes of a mismatch and thereby contribute to a more complete awareness of the different factors that might elicit a mismatch.

Study 1: Pilot study

Methods

Participants

The data of the pilot study are derived from a previous data collection on fear conditioning mechanisms in PTSD patients with borderline personality disorder (Stoffels, Nijs, Spinhoven, Mesbah & Hagenaars, 2016). Permission to use the data for these secondary analyses has been granted by the researchers. The participants with borderline personality disorder and PTSD were recruited at Jelgersma clinic in Oegstgeest, a Dutch clinic specialised in diagnosis and treatment of personality disorders. The sample includes 31 participants (2 males, 29 females) with ages ranging from 19 to 50 with a mean age of 29,52 (SD = 8,37).

Materials

The participants obtained a diagnosis out of two diagnostic methods. First, a clinical diagnosis based on an unstructured clinical interview at intake. The intake consists of two interviews. The first one with a junior and the second one with a senior. Second, a diagnosis based on a cut-point derived from the Dutch version of the PTSD Symptom Scale- Selfreported version (PSS-SR; Arntsz, 1999). The PSS-SR is a self-report questionnaire in which participants have to indicate the extent to which certain symptoms were present during the past week on a 4-point scale, ranging from 0 = never to 3 = five times or more (Engelhard et al., 2007). It measures the severity of present PTSD symptoms. The questionnaire contains 23 items divided over three subscales: re-experiencing (5 items), avoidance (7 items) and arousal (5 items) (Foa, Cashman, Jaycox & Perry, 1997). The re-experiencing subscale contains the core PTSD symptoms and the avoidance and arousal subscales are a.o. containing accompanying anxiety and depression symptoms. Seventeen items are used to assess the total score, these seventeen items each correspond to one of the DSM-IV criteria for PTSD (American Psychiatric Association, 2013). Table 11, admitted in Appendix A, shows all of the PSS-SR items and the seventeen items that contribute to the cutoff score and to which subscale each of the items belongs. The six other items concern experienced feelings related to the traumatic event, and not indicative of the severity of the PTSD symptoms. Wolhfarth and colleagues (2003) looked into the PSS-SR's optimal cutoff score. They have discovered that with a cutoff score of 15, the sensitivity and specificity are optimal (respectively .90 and .90) in predicting DSM-IV PTSD diagnosis. Therefore, this cutoff score of 15 is used in the current study as well. Engelhard and colleagues (2007) conducted a study with 65 PTSD patients, 40 patients with other anxiety disorders and 40 healthy controls. The PSS-SR has been shown to accurately distinguish PTSD patients from healthy controls (5% of the healthy controls were diagnosed with PTSD, and 86% of the PTSD patients were diagnosed with PTSD). As mentioned before, the PSS-SR has some inaccuracy in separating PTSD from other anxiety disorders. 43% of the patients with other anxiety disorders, met the PTSD criteria (Engelhard et al.,

2007). Both the reliability and validity are satisfactory according to the COTAN-guidelines, with a Cronbach's alpha of respectively .91 and .96 (COTAN, 1999; Carlier, Uchelen, Lamberts & Gersons, 1996; Foa et al., 1993).

The Dutch version of the 40 item State-trait anxiety inventory (STAI; self-report) was used to measure the patient's anxiety severity as a current emotional state and as a personality trait (Spielberger, Sydeman, Owen, Marsh & Maruish, 1999; Van der Ploeg, 1981). The STAI consists of two separate questionnaires. *State* anxiety is measured by 20 items; respondents are asked to indicate how they feel about themselves *at the time of assessment* on a 4-point scale, ranging from 1 = not at all to 4 = very much so. *Trait* anxiety is also measured by 20 items rated on a 4-point scale, ranging from 1 = almost never to 4 = almost always. On these items, participants are asked to choose the statement that most closely describes how they *generally* feel (Cao & Liu, 2015; Kabacoff, Segal, Hersen & Van Hasselt, 1997). The Dutch translation of the STAI (STAI-DY) has satisfying psychometric properties. The STAI-trait and state both have a reliability of .91 - .93 (Cronbach's alpha) and the STAI trait has a test-retest reliability of .82 (Ploeg, Defares & Spielberger, 1980).

Analyses

All data are analysed with IBM statistics SPSS 22 (IBM Corp, 2012).

The degree of mismatch between the PSS-SR and the clinician's diagnoses

To assess the degree of mismatch in diagnoses established between the PSS-SR cutoff and the clinicians, descriptives are computed. Of both measurements, a 'yes' and a 'no' diagnosis are present and therefore the degree of mismatch between the PSS-SR and clinician's diagnoses is explored into both directions (i.e. either clinician's diagnosis present and PSS-SR diagnosis absent or the reverse). The degree of mismatch is expressed as the percentage of patients in whom agreement about the diagnosis is present versus absent (i.e.: match versus mismatch). Further, Cohen's Kappa (Cohen, 1968) is calculated. Cohen's Kappa comprise a measure of (in this case) agreement on PTSD diagnosis between two diagnostic measurements of a patient in which a correction is taken into account for chance agreement (Viera & Garrett, 2005). A Cohen's Kappa of 0 is what would be expected as 100% chance (Viera & Garrett, 2005). As mentioned, the guidelines are 0 = no agreement, 0-0.2 = slight, 0.21-0.40 = fair, 0.41-0.60 = moderate, 0.61-0.80 = substantial, and 0.81-1.00 = almost perfect agreement (Landis & Koch, 1977).

Severity of anxiety symptoms as a possible explanation of the degree of mismatch

To investigate whether high scores on the STAI state and trait subscales are predictive of the chance of a mismatch between above cutoff scores of the PSS-SR and the clinician's diagnoses, regression analyses are performed. Due to the dichotomous nature of the outcome variable (mismatch versus match) a logistic regression is used. The independent variable is the score on the STAI trait and state

(continuous) and the dependent variable presence or absence of match (yes/no). The 'Enter' method is used with the variable 'STAI trait' in Box 1 and the variable 'STAI state' in Box 2. The assumptions of the logistic regression (linearity and interdependence) are not violated.

Lastly, the relationship between anxiety severity and both PSS-SR total and subscale severity scores and PSS-SR above cutoff diagnoses are investigated. Bivariate correlations are performed between total scores on the STAI trait and state scale and total PSS-SR and the three subscales of the PSS-SR, using two-tailed Pearson correlations.

Results pilot study

The degree of mismatch between the PSS-SR above cut-off diagnoses and the clinicians' diagnoses

The proportion of patients who gained a PTSD diagnosis or no PTSD diagnosis, established by both the clinician and the PSS-SR self-report cutoff, are displayed in Table 1. Of the total of 28 patients, 12 patients with PTSD and 6 patients without PTSD had a full agreement between the clinician and the PSS-SR above cut-off diagnosis, which means there is full agreement in 18 of the 28 cases (64%). This corresponds to a Cohen's Kappa of .267, which means a fair agreement (Landis & Koch, 1977). The PSS-SR showed a positive diagnosis in more patients (19) than the clinician's diagnosis (15).

Table 1

Number and percentage (%) of patients diagnosed with PTSD or not diagnosed with PTSD by the clinician and the PSS-SR.

	PSS-SR		
Clinician	PTSD (%)	No PTSD (%)	Total
PTSD (%)	12 (43)	3 (11)	15 (54)
No PTSD (%)	7 (25)	6 (21)	13 (46)
Total	19 (68)	9 (32)	28 (100)

Anxiety symptom severity as a possible explanation to explain mismatch (PSS-SR above cutoff scores vs. clinician's diagnosis)

Logistic regression analyses

STAI *state* total scores did not significantly predict whether there is a mismatch or not, $b = -.08$, Wald X^2 (1) = 2.48, $p = .116$. STAI *trait* total scores did significantly predict whether there is a mismatch or not, $b = 0.14$, Wald X^2 (1) = 4.14, $p = .042$. The odds ratio tells us that as the total score on the STAI trait increases by a unit, the change in the odds of a mismatch (rather than a match) is 1.151 (95% CI

[1.005 ; 1.319]): the higher the total score of someone on the STAI trait, the more likely that there is a mismatch. An odds ratio is a measure of association between an exposure and an outcome. The odds ratio represents the odds that an outcome (match or mismatch) will occur given a particular exposure (total score on the STAI trait and state), compared to the odds of the outcome occurring in the absence of that exposure (Szumilas, 2010). Due to the small sample size it was not possible to make two new variables to further explore the directions of the mismatch.

Correlation analyses (Table 2).

Both STAI state and trait scales significantly correlated with the arousal subscale of the PSS-SR, indicating respectively a weak and moderate strength of the correlations (Field, 2013). Further, STAI state correlates significantly with the total PSS-SR (weak correlation) and the PSS-SR subscale avoidance (weak). Interestingly, no significant correlations were found between STAI state and trait and the re-experiencing subscale of the PSS-SR.

Table 2

Correlation matrix of the STAI state and STAI trait questionnaire with the PSS-SR total and the three subscales (Re-experiencing RE - Avoidance AV - Arousal AR).

	PSS-SR total	PSS-SR RE	PSS-SR AV	PSS-SR AR
STAI state	.488**	.321	.395*	.588**
STAI trait	.355	.172	.329	.431*

** Correlation is significant at the .01 level

* Correlation is significant at the .05 level

Conclusions pilot study

The pilot study shows that the agreement between the clinicians' diagnoses and the PSS-SR above cutoff diagnoses of PTSD is only fair in this sample, with higher proportions of patients scoring above cut point on the PSS-SR self-report than on clinician's diagnoses. There are also indications that, with increasing trait anxiety symptom severity, the odds of a mismatch between clinician's diagnoses and the PSS-SR increase. Correlations were found between the STAI state and PSS-SR total, avoidance and arousal but not the re-experiencing subscale, and between STAI trait and the arousal subscale, suggesting symptom overlap between these scales. As a result, patients with higher anxiety symptoms seem to be more likely to be diagnosed with PTSD according to the PSS-SR.

Taking into account the only fair agreement between diagnoses based on different sources of

15

information, this pilot-study emphasizes the importance of investigating the mismatch issue on a larger scale in the second study, evaluating whether we can replicate the relatively over diagnosing of PTSD compared to clinician's diagnoses. The relationship with anxiety symptoms is worthwhile to further explore in the second study as well.

Study 2: the AAA (Altrecht Academic Anxiety center) mismatch study

Data source and participants

The data for the AAA mismatch study were derived from the Altrecht Academic Anxiety outpatient center (AAA), a tertiary mental health institution in Utrecht for diagnosing and treatment of anxiety disorders. At intake, all patients get a clinical unstructured interview to establish diagnoses followed by a SCID-I diagnostic interview by an independent research assistant from the AAA within one week time. Subsequently, based on the outcome of the SCID-I both general and disorder-specific self-report questionnaires are filled in through an internet-based link, including the PSS-SR, Beck depression Inventory (BDI-II) (Beck, Steer & Brown, 1996) and Beck anxiety Inventory (BAI) (Beck & Steer, 1990). From January 2007 to June 2015, 314 newly presented patients received the diagnosis of PTSD according to the SCID-I. Of whom 266 received PTSD as their primary diagnosis and 48 as their additional diagnosis. Those participants that were selected for the current study had a PTSD as a primary diagnosis as a result of the SCID-I interview that was established on them. In other words: the outcome of the SCID-I interview was the starting point for the present study. In 184 patients, data of the PSS-SR were present as well. These patients were selected for the current study.

In sum, this study included 184 participants, 143 female and 41 male. Their age ranged from 18 to 58 years, with a mean age of 34 (SD=9.74). Of all patients, 109 were native Dutch, 70 were migrants and five of unknown origin.

Not all of the instruments or variables (BAI, BDI-II, patient and clinician characteristics) were present for each of the 184 patients. To achieve the largest sample as possible and the highest reliability of the outcomes, it is chosen to maintain the data of all of the 184 patients. Table 3 illustrates the different instruments/variables and the number of patients for whom data was available. Ethical approval was obtained by the Commission on Scientific research of Altrecht. Approval of the medical ethical committee was not necessary because this is not applicable to this study.

Table 3

The different instruments and variables and the number of patients (N) in whom data was available.

Instrument or variable	Available data (N, %)
SCID-I	184, 100%
PSS-SR	184, 100%
Clinicians diagnosis	170, 92%
BDI-II	181, 98%
BAI	83, 45%
Profession of the intaker	165, 89%
Years of experience of the intaker	131, 71%

Methods

The amount of mismatch between PTSD diagnoses is based on discrepancies between diagnoses derived from the following methods: (1) Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders Axis I (SCID-I), (2) the Post-traumatic stress disorder Symptom Scale, self-reported version (PSS-SR) and (3) clinical diagnoses obtained by the clinicians during intake. As the SCID-I PTSD diagnosis was the golden standard for patient inclusion, and therefore no patients were included without a SCID-I PTSD diagnosis, Cohen's Kappa can only be calculated for the discrepancy between the PSS-SR and the clinician's diagnoses because the use of Cohen's Kappa requires both the variables to have the possibility of a 'yes' and a 'no'. The SCID-I only yields a 'yes' on the occurrence of PTSD and is thus not suitable for a Cohen's Kappa calculation. Therefore, the agreement between the PSS-SR and SCID-I and between the clinicians' diagnoses and the SCID-I will be explored by calculating frequencies and expressed in proportions of patients meeting a match. It is expected that Cohen's Kappa between clinician's diagnoses and PSS-SR above cut-off scorers will approach the Cohen's Kappa found in Study 1 (.267 - fair).

Besides a further exploration of the degree of mismatch, attention is paid to the possible explanations of mismatches, at the level of the patient, i.e. whether severity of comorbid anxiety or depressive symptoms contributes to mismatch. The relationship between the occurrence of a mismatch and depression and anxiety severity scores will be explored as well as the correlations between the BDI-II, BAI and the PSS-SR and its subscales. Secondly, explanations will be sought at the level of the clinician, using the following two variables: (1) professional background (medical versus

psychological; levels of education of psychologists) of the clinician as a proxy of educational level, and (2) years of clinical experience of the clinician..

Materials

   SCID-I - In this study, the participants had filled in the Dutch clinical version of the SCID-I (Van Groenestijn, Akkerhuis, Kupka, Schneider & Nolen, 1999). The SCID-I is a structured clinical interview that derives DSM IV axis I diagnoses and needs some training to enable proper use. All research assistants at the AAA are trained in application of the SCID-I diagnostic interview. The structured interview objectifies data obtained from a clinical interview, moreover adding all available clinical information to make a diagnosis. It is a protocol which assists in the formulation of a axis I DSM-IV diagnoses (Steiner et al., 1995). The SCID-I consists of ten modules, namely: mood episodes; psychological and related symptoms; psychotic disorders; mood disorders; substance use; anxiety disorders; somatoform disorders; eating disorders; adjustment disorders and optional disorders (such as, for example, acute stress disorder, or hypomanic episode) (Trimbos Instituut, 2009). The research of Lobbestael, Leurgans & Arntz (2011) showed an substantial inter-rater reliability ($\alpha = .77$) on the diagnosis of PTSD.

   BDI-II - A Dutch version of the Beck Depression Inventory, second version (BDI-II; Van der Does, 2002) was used to measure the severity of the patient's depressive symptoms. The BDI-II is a 21-item self-report measure. Each item is rated on a 4-point scale, yielding summary scores that range from 0 to 63 (Dozois, Dobson & Ahnberg, 1998). The psychometric properties of the BDI-II have been evaluated as satisfactory to good, with an internal consistency of .92 (Cronbach's alpha) and test-retest reliability of .82 (correlation) (Van der Does, 2002). An explanation for the suboptimal content validity might be the overlap with anxiety and neuroticism symptoms (Trimbos Instituut, 2005).

   PSS-SR - The content and psychometric qualities of the PSS-SR are already discussed in the materials section of Study 1.

   BAI - The Dutch version of the Beck Anxiety Inventory (BAI) (Beck & Steer, 1990) was used to measure anxiety severity and entails a 21 item self-report questionnaire, which can be used to measure anxiety severity in several anxiety disorders (Muntingh et al., 2011; Beck & Steer, 1990). The questionnaire consists of physiological and reported anxiety and avoidance symptoms. The respondents are asked to estimate severe each symptom was experienced during the past week, on a scale ranging from 0 (not at all) to 3 (a lot). The BAI has, according to Beck, Epstein, Brown and Steer (1988), a high internal consistency ($\alpha = .92$) and a good test-retest reliability over one week ($r = .75$).

Procedure

The Altrecht dataset provided all the initial measurements (the measurements prior to treatment). The

primary diagnosis based on the SCID-I is PTSD in all of the patients. Other diagnoses based on the SCID-I outcomes are listed as well, ranging from a present PTSD diagnosis to six or more present comorbid diagnoses. To assess the clinician's diagnoses, the patient numbers were matched to their corresponding files in the EPD (Elektronisch Patiënten Dossier - a digital community for assessing patient files). In these files, the required information was sought in the intakefile. The intake information yields the initial diagnoses given by the intaker, the profession of the intaker (i.e. psychotherapist, psychiatric nurse, psychiatrist, et cetera) and the name of the intaker. The intakers name was necessary to later obtain the number of years of experience at the time of the intake.

Analysis

The data were analysed with IBM statistics SPSS 22 (IBM Corp, 2012).

Degree of mismatch between SCID-I, PSS-SR and clinician's diagnosis

To assess the agreement between the clinician's diagnosis and the PSS-SR, Cohen's Kappas are calculated. To assess the agreement between the SCID-I and the PSS-SR and the clinician's diagnosis, descriptives (i.e. frequencies and percentages) will be used. Descriptives will also be used to see which method diagnoses PTSD most frequently. Cohen's Kappa have not been calculated for the agreement between the SCID-I and the PSS-SR.

Patient factors as possible explanations for mismatch

To investigate whether the scores on the BAI and BDI-II are predictive of presence of a possible mismatch (PSS-SR vs. clinician), a regression analysis is performed. Due to the dichotomous nature of the outcome variable (diagnosis: PTSD present versus PTSD not present) a logistic regression analysis is carried out. The independent variables are the scores on the BAI and BDI-II (continuous) and the dependent variable the presence or absence of match (yes/no). The assumptions of the logistic regression (linearity and interdependence) are not violated. The 'Enter' method is used with variable 'BDI-II total score' put in Box 1 and 'BAI total score' in Box 2. Moreover, bivariate correlations are computed to see whether the scores on the BAI and BDI-II correlate with the outcome on the total PSS-SR and the three subscales of the PSS-SR, using two-tailed Pearson's correlations. This is done to explore whether there might be a symptom overlap in the items of the BAI and BDI-II with the total PSS-SR and its subscales. Finally, to give more insight into the mismatch issue, an overview of all the diagnoses given by the clinicians when they missed the diagnosis of PTSD will be given.

Clinician's factors as a possible explanation for the mismatch

To investigate whether the experience of the clinician in years is associated with the degree of mismatch (clinician's diagnosis vs. SCID-I) (yes/no), Chi-square is computed Only mismatch relationships can be explored in which SCID-I PTSD diagnosis is present and clinician's PTSD diagnosis is absent. The clinicians are divided in in six groups according to their years of experience,

to obtain balanced numbers per group, a requirement needed to perform logistic regression analyses (Table 4).

To test whether education level of the clinician contributes to the degree of mismatch (clinician's diagnosis vs. SCID-I) (yes/no), Chi-square is computed. Similar to the variable experience in years, only mismatch relationships are explored in which SCID-I PTSD diagnosis is present and clinician's PTSD diagnosis is absent. Clinicians were divided into five groups, according to their education level (Table 5). Psychiatric nurses and medical interns have a medical background, whereas the others have a psychological background (by training). Besides, the groups differ in education level and years of education. The chi-square analysis shows whether the degree of mismatch differs significantly between the different educational levels.

Table 4

Years of experience of clinicians divided in six groups, including number of diagnoses per group.

	Years of clinical experience					
	0-1	1-3	3-5	5-10	10-20	20>
Number of clinicians (N)	18	25	19	18	27	24

Table 5

Number of diagnosing clinicians cross five educational groups (as a proxy for education level).

	Psychiatric nurses	Medical interns	Basic Psychologists	Health-care psychologists	Psychotherapists and clinical psychologists
Number of clinicians (N)	26	27	36	43	33

Results

Degree of mismatch between SCID-I, PSS-SR and clinician's diagnosis

To assess the degree of mismatch between the SCID-I on the one hand, and the PSS-SR above cutoff diagnoses and clinician's diagnoses on the other, descriptives were used. Table 6 shows the number of patients in whom a mismatch of diagnoses was found between SCID-I PTSD diagnoses one the one

hand and both PSS-SR above cut-off and clinician's diagnoses on the other. The degree of match between the PSS-SR and the SCID-I is 92%, indicating that in 14 out of 184 patients the PSS-SR did not indicate a PTSD diagnosis whereas the SCID-I did. Regarding the match between SCID-I and clinician's diagnosis, the degree of mismatch and match is 31% and 69%, respectively, indicating that in 53 (out of 184) patients the clinician did not establish a PTSD diagnosis whereas SCID-I did establish a diagnosis.

To further explore the degree of mismatch between the PSS-SR and clinician's diagnoses into both directions (i.e. either clinician's diagnosis present and PSS-SR diagnosis absent or the reverse), Cohen's Kappas were calculated. Table 7 shows the number of patients in whom number of patients in whom mismatch occurred between clinician's PTSD diagnosis and PSS-SR above cutoff diagnosis. Interestingly, there is an agreement between clinician's diagnoses and PSS-SR in 115 (out of 170) cases (68%) and there are 55 (out of 170) cases of mismatch (32%), with similar numbers of mismatch into both directions, i.e the clinician establishing a PTSD diagnosis but not the PSS-SR or the reverse. This corresponds to a Cohen's Kappa of .056, which means there is 'no agreement' between PSS-SR and clinician's diagnoses (Landis & Koch, 1977). The PSS-SR above cut-off self-report diagnosed significantly more patients with PTSD (156) than did the clinicians' interviews (117). In 8 cases the clinician diagnosed PTSD where the PSS-SR above cutoff did not.

Table 6

Number of patients with matching/ mismatching diagnosis between SCID-I PTSD diagnosis present on the one hand and the PSS-SR cut-point derived PTSD diagnosis and clinician's diagnosis on the other.

	Diagnostic method				
	PSS-SR		Clinician's diagnosis		
	No match (%)	Match (%)	No Match (%)	Match (%)	Missing CD
SCID-I Diagnosis PTSD	14 (8)	156 (92)	53 (31)	117 (69)	14

Table 7

Number of patients who matched/ mismatched between PSS-SR PTSD diagnosis and clinician's PTSD diagnosis.

	PSS-SR		
Clinician	PTSD (%)	No PTSD (%)	Total
PTSD (%)	109 (64)	8 (5)	117
No PTSD (%)	47 (28)	6 (3)	53
Total	156 (99)	14 (1)	170

Patient factors as a possible explanations for mismatch

Logistic regression on PSS-SR above cutoff diagnosis versus clinician's diagnosis

The total scores on the BDI-II did not significantly predict whether there is a mismatch between the clinician's diagnosis and the PSS-SR diagnosis or not, $b = .012$, Wald X^2 (1) = 0.253, $p = .615$. The total score on the BAI did not significantly predict whether there is a mismatch between the clinician's diagnosis and the PSS-SR diagnosis, or not either, $b = -.010$, Wald X^2 (1) = 0.258, $p = .612$.

Due to the contradictory outcomes in comparison to the literature and the pilot study, two additional exploratory logistic regressions were carried out. Two new categorical variables were computed. The first consisting of 0 = match between the PSS-SR and clinician and 1 = mismatch with the clinician not diagnosing PTSD and the PSS-SR diagnosing PTSD. The second variable consisted of 0 = match and 2 = mismatch with the clinician diagnosing PTSD and the PSS-SR not diagnosing PTSD.

Match versus mismatch with the PSS-SR diagnosing PTSD and the clinician not diagnosing PTSD (0 vs. 1). The total score on the BAI did not significantly predict whether there is a mismatch or not, $b = -.014$, Wald X^2 (1) = 0.462, $p = .497$. BDI-II total scores did not either significantly predict whether there is a mismatch or not, $b = .050$, Wald X^2 (1) = 3.399, $p = .065$. .

Match versus mismatch with the clinician diagnosing PTSD and the PSS-SR not diagnosing PTSD (0 vs. 2). The total score on the BAI did not significantly predict whether there is a mismatch or not, $b = .073$, Wald X^2 (1) = .538, $p = .463$. BDI-II total scores did significantly predict whether there is a mismatch or not, $b = -.407$, Wald X^2 (1) = 4.421, $p = .036$. The odds ratio can be interpreted as follows: as the total score on the BDI-II increases by a unit, the change in the odds of a mismatch (rather than a match) is 0.665 (95% CI [0.455 ; 0.971]): the higher the total score of someone on the

BDI-II, the less likely it was that the clinician would diagnosed PTSD and the PSS-SR above cutpoint would not.

Correlations

Correlations were calculated to explore the relationships between the BDI-II, the BAI and the total PSS-SR and subscale scores of the PSS-SR. The BDI-II, BAI, PSS-SR and its subscales are added to the correlation matrix (Table 8). The BDI-II and BAI both correlate significantly on the $\alpha = 01$ level with all of the subscales and with the total PSS-SR, with moderately to high correlations (Field, 2013).

Table 8

Correlation matrix of the total BDI-II and BAI with the PSS-SR total and the three subscales (Re-experience - Avoidance - Arousal)

	PSS-SR total	PSS-SR RE	PSS-SR AV	PSS-SR AR
BDI-II	.736**	.533**	.678**	.645**
BAI	.602**	.496**	.466**	.615**

** Correlation is significant at the .01 level

Further exploratory findings: Diagnoses given by the clinicians who did not diagnose PTSD whereas SCID-I had established a PTSD diagnosis as primary diagnosis

Table 9 shows the most frequent diagnoses given by the clinicians in those patients with SCID-I PTSD in whom clinicians had not established the diagnosis of PTSD (n=53). The clinicians often diagnosed several disorders, hence the amount of diagnoses extending 53 over all 184 patients. To maintain a comprehensible table and because of relevance, only the most frequent diagnoses (>3 times) are included in the table. Appendix B contains the extended table (Table 12) with all of the diagnoses. Moreover, it is explored whether these diagnoses are consistent with the additional SCID-I diagnoses (Table 9). As Table 9 shows, when the clinicians did not establish the diagnosis of PTSD, they instead mostly diagnosed a depressive disorder, OCD, anxiety disorder NAO, acute stress disorder, agoraphobia, generalized anxiety disorder or a social anxiety disorder. In 59% of the cases are these diagnoses consistent with the comorbid diagnoses on the SCID-I. Following this, 41% of the given diagnoses by the clinicians are not consistent with the co-morbid diagnoses on the SCID-I, which means that the degree of match between SCID-I diagnoses and clinician's diagnoses was particularly low in these patients; not only did the clinicians not establish a PTSD diagnosis, they diagnosed a

disorder which was not diagnosed by the SCID-I.

Table 9

Number of patients diagnosed with alternative primary diagnoses from PTSD (according to the SCID-I) and the correspondence to the SCID-I diagnoses (besides PTSD). When the clinicians did not establish the diagnosis of PTSD, they instead mostly diagnosed a depressive disorder, OCD, anxiety disorder NAO, acute stress disorder, agoraphobia, generalized anxiety disorder or a social anxiety disorder.

Diagnosis by clinician	Number of times diagnosed by the clinician (N)	Consistent with the SCID-I?
No diagnosis	6	Yes: 3 No: 3
Depression	21	Yes: 18 No: 3
Obsessive compulsive disorder (OCD)	6	Yes: 2 No: 4
Anxiety disorder NAO	5	Yes: 0 No: 5
Acute stress disorder	4	Yes: 0 No: 4
Panic disorder without agoraphobia	4	Yes: 3 No:1
Panic disorder with agoraphobia	4	Yes: 4 No: 0
Generalized anxiety disorder (GAS)	4	Yes: 2 No: 2
Social anxiety disorder	4	Yes: 2 No: 2

Clinicians' factors as possible explanations for mismatch

To test whether the degree of mismatch (clinician's diagnosis vs. SCID-I) is influenced by the education level and educational background of the clinicians, Chi-squares have been calculated (profession X mismatch). The degree of mismatch did not differ by the clinician's profession, X^2 (4, N = 165) = 2.428, p = .658.

To test whether the degrees of mismatch (clinician's diagnosis vs. SCID-I) are influenced by the clinical experience of the clinicians in years Chi-squares have been calculated as well. The degree of mismatch did significantly differ by the number of years of clinical experience, X^2 (5, N = 131) = 15.750, p = .008. Table 9 shows that clinicians who have 0-1 and 20 > years of experience have the highest degree of mismatch between clinicians' and SCID-I diagnosis, 50% and 42% respectively.

Table 10

Degree of mismatch (clinician versus PTSD) by years of clinical experience described in percentages

	Experience clinician in years					
	0-1	1-3	3-5	5-10	10-20	20 >
Match (%)	50	84	84	67	92	58
Mismatch (%)	50	16	16	33	8	42

Discussion

As mentioned, the presence of PTSD is accompanied by considerable suffering, functional impairment and high costs for the community and government. This research project has attempted to provide more insight into the mismatch problem between standardized diagnoses, clinicians' diagnoses and self-report derived diagnoses, and the agreement (hypothesis 2-1), or amount of match and mismatch, between the SCID-I and the PSS-SR; between the SCID-I and the clinician's diagnosis; and between the clinician's diagnosis and the PTSS-SR was explored. Moreover, explanations for the mismatches were explored at the level of the patient (2-2a) and the level of the clinician (2-2b). Prior to current research, a number of expectations were drawn. Below, the different questions, their corresponding expectations and outcomes will be discussed. The outcomes of the pilot study were already presented under 'Pilot study - Conclusion' and will now be integrated into the outcomes of Study 2. Moreover, a general discussion/conclusion, implications for the clinical practice and future research directions will be presented.

The amount of (mis)match between the several diagnostic instruments

It was expected that the mismatch between the SCID-I and the clinicians would be fair to moderate. Current results show a 69% match and 31% mismatch. As a PTSD diagnosis on the SCID-I served as the inclusion to the study, this means that the clinicians did not diagnose 31% of the PTSD cases that were diagnosed according to the SCID-I interview. Although we were unable to calculate Cohen's Kappas, it is likely that 31% mismatch also indicates "no" or at best "poor" agreement'. The outcome is therefore lower than expected. However, one should be cautious about this negative result. Rettew and colleagues (2009) found a higher proportion of agreement (moderate) than the current study. Again, comparison is difficult since Rettew and colleagues (2009) did not mention percentages of agreement, but calculated Cohen's Kappa in contrast to our study. Further, Rettew and colleagues (2009) included 888 subjects in their meta-analyses, which can provide more reliable results. However, our results are in line with Rettew and colleagues (2009) in that the clinicians reported less PTSD than the SDI interviews.

Concerning the agreement between the PSS-SR and the SCID-I, it was expected that the agreement would be moderate and would approach the percentages found in the aforementioned studies. With an agreement of 92% found in the current study, this expectation has been approximately met. The agreement found in our study is slightly higher than found by Engelhard and colleagues (2007), Wohlfarth and colleagues (2003) and Foa and colleagues (1993), who found 75%, 85% and 86% agreement, respectively. This difference might be explained by the different DSM criteria which are used for the SCID-I and PSS-SR. The studies all used the DSM-III criteria instead of the DSM-IV criteria used in the current study. Speculating, the upgrade to DSM-IV criteria might have naturally led to a more accurate PTSD diagnosis that was more in line with the criteria measured

with the PSS-SR. The fairly low mismatch could possibly be explained by the fact that the SCID-I questions and PSS-SR items are both based on DSM-IV criteria and therefore resemble each other. Moreover, Wohlfarth used the CIDI instead of the SCID-I, which might lead to different outcomes as well. The CIDI seems to be a more sensitive instrument to attain a diagnosis than the SCID-I, indicating that the CIDI might slightly overreport in comparison to the SCID-I. Wohlfarth and colleagues (2003) have found a moderate Cohen's Kappa at an agreement of 85%. With caution, one could say that the agreement found in the current study is at least moderate. This outcome is satisfactory. However, there are also differences between the instruments in the way they obtain a diagnosis. The PSS-SR obtains a diagnosis out of a cumulative total score with a cut-off point, and is based on self-report, i.e. on what patients actually think of their signs and symptoms themselves. There are no items or (core)symptoms especially necessary to diagnose PTSD. The SCID-I however, uses the composition of the DSM-IV criteria and pursues these by questioning all the core symptoms. PTSD can only be diagnosed when the interviewer judges that the person possesses all of these symptoms. When the focus is not only on the presence of core symptoms, as is the case with the PSS-SR, someone could attain the diagnosis of PTSD by scoring high on the other symptoms/items. The PSS-SR picks up anxiety and depressive symptoms as well (Engelhard et al., 2007; Meltzer et al., 2012). The remaining mismatch could possibly explained by this atypicality of the PSS-SR.

No expectations could be formulated according to the mismatch between the PSS-SR and clinician, due to the absence of literature. The pilot study showed a fair agreement. It was likely that the agreement in study 2 would be similar to the pilot outcomes. However, the agreement between the PSS-SR and the clinician's diagnosis is even lower in study 2 than in the pilot study. A mismatch of 32% was found, which means there was no agreement. The PSS-SR reports PTSD substantially more than the clinician, which is in line with the pilot study.

In sum, the clinician shows the lowest reporting of a PTSD diagnosis, compared to the SCID-I and the PSS-SR. What could be reasons for these discrepancies? One explanation could be that during an interview with a patient clinicians perform open interviews, that are not necessarily based on rules or diagnostic criteria and they rely on heuristic principles to find a diagnosis (Garb, 2005; Kim & Ahn, 2002; Tversky & Kahneman, 1974). Moreover, clinicians are highly influenced by the training background including their psychotherapeutic background. For instance, a family therapist who performs the intake will preferentially focus on family circumstances that might had led to the current psychological state of the patient. As a result: clinicians might skip or forget about necessary steps in the diagnostic process (Garb, 2005). Besides the clinician, attention should be paid to the PSS-SR and its possible weaknesses. Self-report measures are prone to biases, they are highly influenced by how the person is feeling at the time of filling in the questionnaire. Moreover, they are influenced by sociability traits of the respondent, i.e. the tendency to answer positively to the items that are provided.

At least, the advantage of structured interviewing is that all possible diagnoses are systematically addressed, leaving less room for "random" questioning as seen in clinical intakes.

Explanations for the mismatch at the patient's level

Concerning the explanations sought at the level of the patients, it was expected that the presence of comorbid anxiety and/or depressive symptoms would lower the agreement between interviews and self-report because these symptoms would obscure PTSD diagnosis both on the PSS-SR and as established by clinicians. Study 1 indeed showed a relationship between lifetime (trait) anxiety and the occurrence of a mismatch between the PSS-SR and the clinician. The higher the score on the STAI trait, the greater the chance of a mismatch occurring. However, the directions of the mismatch itself remained unclear. Study 1 also showed that the STAI correlates with the total PSS-SR and especially, with the non PTSD specific arousal and avoidance subscales of the PSS-SR, suggesting blurring of a PTSD self-report diagnosis by symptom overlap with more general anxiety symptoms. As Engelhard and colleagues (2007) mentioned, the hyperarousal and avoidance symptoms are both occurring in PTSD as well as in anxiety disorders.

Contradictory to the literature and the pilot study, the larger scale study showed that BAI total scores were not associated with the chances of a mismatch occurring. However, this lack of finding might be due to power issues, as a relatively low number of participants filled in the BAI. BDI-II total scores (with high numbers of patients having filled in the BDI-II) seemed to be associated with odds of mismatch; there is the possibility that a higher score on the BDI-II is associated with a higher mismatch. This outcome is according to the expectations. As Engelhard and colleagues (2007) mentioned, depressive symptoms are also measured by the PSS-SR because of symptom overlap between PTSD, depression and a depressed mood originated from an anxiety disorder. This might lead to an increase in false positives on the PSS-SR. On the other hand, at intake when someone suffers from many depressive symptoms, this manifests itself more easily to the clinician, not only by interview but for instance also by an psychomotor symptoms of inhibition, possibly obscuring an underlying diagnosis of PTSD. As Meltzer and colleagues (2012) mentioned, physicians are more likely to recognize these depressive symptoms, to ignore the underlying PTSD and to mislabel PTSD patients as having a depression as their primary diagnosis.

In Study 2, the BDI-II and BAI both correlated significantly with all of the PSS-SR subscales and with the total PSS-SR. This is mostly in agreement with Study 1 and the current literature. It seems striking that the BDI-II and BAI were both associated with the re-experiencing subscale, as this subscale contains PTSD specific symptoms. However, Brewin (2003) mentioned that even though intrusive memories and flashbacks - as represented by the re-experiencing subscale - are seen as distinctive PTSD symptoms, these can also be present in panic disorder, social phobia or a depression. These results all indicate that the PSS-SR is influenced by self-reported anxiety and depressive

symptoms and therefore stands the risk of diagnosing someone with PTSD when the disorder is not present. However, the high match between SCID diagnoses and PSS-SR argue against this.

It is noteworthy and worrying that clinicians completely missed the diagnosis of PTSD and that 41% of those clinicians diagnosed a disorder which was not even present according to the SCID-I. Mostly, the clinicians diagnosed a depression, OCD, anxiety disorder NAO, acute stress disorder, panic disorder and generalized anxiety disorder but not a PTSD. It is notable that these disorders are all anxiety or depressive disorders. This indicates that the presence of comorbid anxiety and/or depressive symptoms seems to blur the clinician's diagnosis, leading to underreporting by the clinicians. It blurs the PSS-SR diagnosis as well, but the PSS-SR seems to overreport. This is in consensus with the other outcomes and with Engelhard (2007) and Melzer (2012) and colleagues, who argued that a depression, panic disorder or GAD is often diagnosed instead of PTSD.

Explanations for the mismatch at the clinician level

Regarding the explanations at the level of the clinician, it was expected that there would be a difference in diagnostic accuracy between different professions. It was expected that the more education someone gained, the more accurate their diagnostic skills would be. No expectation was formulated regarding the difference between medical or psychological background, due to absence of literature. The current study shows that there is no difference in diagnostic accuracy between different professions.

Moreover, an effect of clinician's year of experience would play a role in diagnostic accuracy. The results of this study show that clinical experience influences diagnostic accuracy. Notably, clinicians with 0 to 1 years of experience and clinicians with more than 20 years of experience seem to be the least accurate. They missed 50% and 42% of the PTSD diagnosis according to the SCID-I, respectively. A reason that clinicians with 0 to 1 year of experience often fail to be accurate about their diagnosis, could be that they do not yet have the knowledge they need to compose a diagnosis. In other words, too little mastery of content and experience. Beside, novices have little knowledge about how to use specific techniques to elicit the accurate information from the patient to enable a sound diagnosis (Dawes, 1989). On the other hand, the results suggest that clinicians with more than 20 year of diagnostic experience also often fail to be accurate. Experienced clinicians use other reasoning processes than novices (Elstein & Schwarz (2002). They only use a hypothetical-deductive strategy in complex cases. Experienced clinicians are more inclined to use pattern recognition and automatic retrieval for composing a diagnosis (Elstein & Schwarz, 2002). Therefore, experience is helpful in order of knowledge and cleverness. But it could also become a pitfall when they rely too much on their clinical experience.

General discussion/conclusion

Overall, the findings concerning the amount of mismatch are disturbing. It seems that the clinicians

miss most of the PTSD cases. The PSS-SR misses fairly less PTSD cases, with 8% compared to 31% of the clinician. It is a substantial shortcoming of the current study that the SCID-I yielded only PTSD-positives, since over-reporting could not be investigated. Moreover, we should carefully interpret the results of the study regarding the analyses containing the SCID-I, since the SCID-I is used as the golden standard in this study. And despite the excellent psychometric properties, we cannot guarantee that the SCID-I does not generate false positives or false negatives. There are indications, that the PSS-SR overreport and the clinicians underreport. According to the pilot study present anxiety might increase the chance of a mismatch occurring, however, according to Study 2 current anxiety severity does not seem to be associated. When depressive symptoms are present, this results in a higher chance of a mismatch occurring. Depressive symptoms, or a depression, seems to blur both the PSS-SR and the clinician. Moreover, both anxiety and depressive symptoms, as measured by the BDI-II and BAI, show symptom overlap with the PSS-SR and its subscales, indicating blurring as well. No differences in diagnostic accuracy between the different professions are found. In contrast, clinical experience does influences diagnostic accuracy. Clinicians with 0 to 1 years of experience and clinicians with more than 20 years of experience are the least accurate - resulting in the highest mismatch.

Implications for the clinical practice

When taking the results of both studies and the literature into account, it seems important not to base a diagnosis solely on the intake findings. Clinicians might miss symptoms, do not weigh symptoms equally, rely on cognitive heuristics and do not have all the DSM criteria accurately available by themselves. This could, among other things, lead to false diagnoses which causes an unsuited psychotherapy. Therefore, a (semi) structured interview assessed by an independent interviewer should be added to the diagnostic process, to add to the information given by the intaker. One should however be aware that it is possible be that the SCID-I also misses PTSD diagnosis. Moreover, during intake, clinicians should always ask about traumatic experiences in order to prevent that PTSD is seen as another psychiatric disorder. Both anxiety disorders and depressive disorders have symptom overlap with PTSD, but the occurrence of a trauma (including re-experiencing) is a unique key factor of PTSD, it discriminates PTSD from anxiety disorders and depressive disorders.

Another way to decrease mismatch is to make clinicians, especially the more experienced ones, aware of their cognitive biases and heuristics and explain how these fallacies influence their diagnostic accuracy. And more attention should be paid to novices, since their diagnostic accuracy is lower as well. This might be done with for example more training and more supervision.

The PSS-SR is a suitable instrument to assess PTSD severity and to follow patients on treatment success. It is an easy, less incriminating, time-saving and therefore inexpensive way of an initial screening. However, an anxiety disorder or depressive disorder can be falsely seen as a PTSD.

Patients who score positive on PTSD according to the PSS-SR, should therefore always be further examined for PTSD core symptoms like an experienced trauma. Anyway, it is clear that there can never be a decision made on the basis of one single method, but that information based on different sources is required to derive the most reliable diagnoses.

Future research directions

As already mentioned, a substantial shortcoming of the current study is that the SCID-I yielded only PTSD-positive diagnoses and therefore the comparison between SCID-I negative PTSD diagnoses in cases in which clinicians or self-report did suggest a PTSD diagnosis could not be investigated. Future research should use data with SCID-I PTSD-negatives as well. As there are indications that the PSS-SR over-report, this could be further investigated.

It is important to further investigate whether clinical judgment is dependent on a clinician's profession. This should be done in an experimental design including more subjects, to draw firm conclusions. Current study is the first investigating this matter, it could be potentially very important to clinical practice. The proposed techniques to improve (clinical) judgment should be implemented in daily clinical practice. After implementation, one should investigate whether the implementation would lead to improvements in judgment and lower the amount of mismatch. Moreover, the mismatch issue is not solely limited to PTSD. It is worth investigating the issue with other disorders.

At last, the current study obtained data from one single clinic. It is desirable to investigate this issue on a larger scale, with data from different clinics to diminish potential institutional biases.

Epilogue

The current study has addressed a highly interesting and important issue in mental health in the Netherlands, i.e. that most diagnoses and treatments are based on one person intake procedures, but that these procedures are vulnerable to missing diagnoses, apparently. The current project has been a first step to address this important issue.

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3th ed.). Washington, DC: Author

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Andreas, S., Theisen, P., Mestel, R., Koch, U. & Schulz, H. (2009). Validity of routine clinical DSM-IV diagnoses (Axis I/II) in inpatients with mental disorders. *Psychiatry Research, 170,* 252-255.

Arntz, A. (1993). Nederlandse vertaling van de PSS-SR. Universiteit Maastricht: intern rapport.

Beck, A. T., & Steer, R. A. (1990). BAI, Beck anxiety inventory. San Antonio: Psychological Corporation.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. San Antonio: Psychological Corporation.

Beck AT, Epstein N, Brown G, Steer RA, 1988: An inventory for measuring clinical anxiety: psychometric properties. *Journal of Consulting and Clinical Psychology, 56*, 893–897.

Brewin, C. R. (2003). *Posttraumatic stress disorder: Malady or myth?* New Haven: Yale University Press.

Cao, Y., & Liu, Z. K. (2015). Use of the State-Trait Anxiety Inventory with children and adolescents in China: Issues with reaction times. *Social Behavior and Personality: an international journal*, *43*, 397-410.

Carlier, I. V. E., Uchelen, J. J., Lamberts, R. D. & Gersons, B. P. R. (1996). A brief screening test for the diagnosis posttraumatic stress disorder. *Tijdschrift voor Psychiatrie, 38,* 624-629.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, *70*, 213.

Cotan. Cotan documentatie NIP. Consulted at 8 januari 2016, from http://www.cotandocumentatie.nl/test_details.php?id=454.

Dawes, R. M. (1989). Experience and validity of clinical judgement: The illusory correlation. *Behavioral Sciences & the Law, 7,* 457-467.

Dozois, D. J., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory–II. *Psychological Assessment*, *10*, 83.

Elstein, A.S., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *BMJ, 324*, 729-732.

Engelhard, I.M., Arntz, A., & Van den Hout, M.A. (2007). Low specificity of symptoms on the post-traumatic stress disorder (PTSD) symptom scale: A comparison of individuals

with PTSD, individuals with other anxiety disorders and individuals without psychopathology. *British Journal of Clinical Psychology, 46*, 449-456.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics.* Sage Publications Ltd.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1997). *User's guide for the Structured clinical interview for DSM-IV axis I disorders SCID-I: clinician version.* New York: American

    Psychiatric Pub.

Foa, E. B., Riggs, D. S. Dancu, C. V. & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress, 6*, 459-473.

Galea, S., Nandi, A. & Vlahof, D. (2005). The epidemiology of post-traumatic stress disorder after disasters. *Epidemiologic Reviews, 27*, 78-91.

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105*, 387-396.

Garb, H. H., & Boyle, P. A. (2003). Understanding why some clinicians use pseudoscientific methods: Findings from research on clinical judgment. In Lilienfeld, S.O., Lynn, S.J., & Lohr, J.M. (Eds.), Science and pseudoscience in clinical psychology (17-38). New York: Guilford

Garb, H. N. (2005). Clinical judgement and decision making. *Annual Review of Clinical Psychology, 1,* 67-89.

Gregory, R. J. (2007). *Psychological Testing: History, Principles, and Applications* (5th ed.). Boston: Pearson Education.

IBM Corp. (2012). IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.

Jones, K. D. (2010). The unstructured clinical interview. *Journal of Counseling & Development*, *88*, 220-226.

Kabacoff, R. I., Segal, D. L., Hersen, M., & Van Hasselt, V. B. (1997). Psychometric properties and diagnostic utility of the Beck Anxiety Inventory and the State-Trait Anxiety Inventory with older adult psychiatric outpatients. *Journal of anxiety disorders*, *11*, 33-47.

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M. & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity study. *Archives of General Psychiatry, 52,* 1048-1060.

Kim, N. S. & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology, 131,* 451-476.

Kovess, V., Sylla, O., Fournier, L. & Flavigny, V. (1992). Why discrepancies exist between structured diagnostic interviews and clinicians' diagnoses. *Social Psychiatry and Psychiatric Epidemiology, 27,* 185-191.

Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Lobbestael, J., Leurgans, M., & Arntz, A. (2011). Inter□rater reliability of the Structured Clinical Interview for DSM□IV Axis I disorders (SCID I) and Axis II disorders (SCID II). *Clinical Psychology & Psychotherapy*, *18*, 75-79.

Maruish, M.E. (1999). *The use of psychological testing for treatment planning and outcomes assessment* (2nd ed.). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Meltzer, E. C., Averbuch, T, Samet, J. H., Saitz, R., Jabbar, K., Lloyd-Travaglini, C. & Liebschutz, J. M. (2012). Discrepancy in Diagnosis and Treatment of Post-traumatic Stress Disorder (PTSD): Treatment for the Wrong Reason. *The Journal of Behavioral Health Services & Research, 39,* 190-201.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., ... & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American psychologist, 56,* 128.

Miller, P. R., Dasher, R., Collins, R., Griffiths, P. & Brown, F. (2001). Inpatient diagnostic assessments: accuracy of structured vs. unstructured interviews. *Psychiatry Research, 105,* 255-264.

Muntingh, A. D., van der Feltz-Cornelis, C. M., van Marwijk, H. W., Spinhoven, P., Penninx, B. W., & van Balkom, A. J. (2011). Is the beck anxiety inventory a good tool to assess the severity of anxiety? A primary care study in The Netherlands study of depression and anxiety (NESDA). *BMC family practice,12*, 66.

Rettew, D. C., Lynch, A. D., Achenbach, T. M., Dumenci, L. & Ivanova, M. Y. (2009). Meta-analysis of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research, 18,* 169-184.

Sapyta, J., Reimer, M. & Bickman, L. (2005). Feedback to clinicians: Theory, research, and practice. *Journal of Clinical Psychology, 61,* 145-153.

Spengler, P. M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S., ... & Rush, J. D. (2009). The meta-analysis of clinical judgment project effects of experience on judgment accuracy. *The Counseling Psychologist*, *37*, 350-399.

Spielberger, C. D., Sydeman, S. J., Owen, A. E., Marsh, B. J. & Maruish. M. E. (1999). *Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI)*. Mahwah, US: Lawrence Erlbaum Associates Publisher.

Stein, M. B., McQuaid, J. R., Pesrelli, P., Lennox, R. & McCahill, M. E. (2000). Posttraumatic stress disorder in the primary care medical setting. *General Hospital Psychiatry, 22,* 261-269.

Steiner, J. L., Tebes, J. K., Sledge, W. H. & Walker, M. L. (1995). A Comparison of the Structured Clinical Interview for DSM-III-R and Clinical Diagnoses. *The Journal of Nervous and Mental Disease, 183*, 365-369.

Stice, E., Telch, C. F., & Rizvi, S. L. (2000). Development and validation of the Eating Disorder Diagnostic Scale: a brief self-report measure of anorexia, bulimia, and binge-eating disorder. *Psychological assessment, 12*, 123

Stoffels, M., Nijs, M., Spinhoven, P., Mesbah, R., & Hagenaars, M.A. (2016). Emotion Regulation and Fear Bradycardia in Patients with Borderline Personality Disorder and Healthy Controls. Manuscript submitted for publication.

Szumilas, M. (2010). Explaining Odds Ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 19,* 227-229.

Tversky, A. & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Trimbos Instituut. (2005). *Ernstmeting*. Consulted at 8 januari 2016, from http://www.ggzrichtlijnen.nl/index.php?pagina=/richtlijn/item/pagina.php&id=1261&richtlijn_id=88

Trimbos Instituut. (2009). *(Semi)gestructureerde diagnostische interviews*. Consulted at 8 januari 2016, from http://www.ggzrichtlijnen.nl/index.php?pagina=/richtlijn/item/pagina.php&id=1144&richtlijn_id=35

Van Groenestijn, M.A.C., Akkerhuis, G.W., Kupka, R.W., Schneider, N., & Nolen, W.A. (1999). *Gestructureerd klinisch interview voor de vaststelling van DSM-IV as-I stoornissen (SCID-I)* Lisse, The Netherlands: Swets & Zeitlinger.

Van Balkom, A.L.J.M., Van Vliet, I.M., Emmelkamp, P.M.G., Bockting, C.L.H., Spijker, J., Hermens, M.L.M., & Meeuwissen, J.A.C. (2013). *Multidisciplinaire richtlijn Angststoornissen (Derde revisie). Richtlijn voor de diagnostiek, behandeling en begeleiding van volwassen patiënten met een angststoornis*. Utrecht: Trimbos-instituut.

Van der Does, A. J. W. (2002). Handleiding bij de Nederlandse versie van beck depression inventory—second edition (BDI-II-NL). *Lisse, NL: Harcourt Test Publishers*.

Van der Ploeg, H.M. (1981). *Zelf-beoordelings vragenlijst, handleiding*. Lisse, The Netherlands: Swets and Zeitlinger.

Van der Ploeg, H. M., Defares, P. B. & Spielberger, C. D. (1980). *Handleiding bij de Zelf*

Beoordelings Vragenlijst, ZBV. Lisse: Swets & Zeitlinger.

Vermetten, H. G. J. M. (2014). Strijd van binnen. Consulted at 18 november 2015, from
Universiteit Leiden, via https://openaccess.leidenuniv.nl/bitstream/handle/1887/26061/
Oratie%20Vermetten.pdf?sequence=1

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic.
Fam Med, *37*, 360-363.

Williams, J.B.W., Gibbon, M., First, M.B., Spitzer, R.L., Davies, M., Borus, J., … & Wittchen, H. U
(1992). The Structured Clinical Interview for the DSM-III-R (SCID): Part II: multisite test-
retest reliability. *Archives of General Psychiatry, 49*, 630-636.

Wohlfarth, T. D., van den Brink, W., Winkel, F. W. & ter Smitten, M. (2003). Screening for
Posttraumatic Stress Disorder: An Evaluation of Two Self-Report Scales Among
Crime Victims. *Psychological Assessment, 15,* 101-109.

Appendix A - The PSS-SR

Table 11

All of the PSS-SR items and the seventeen items that contribute to the cutoff score(*) and to which subscale they belong: re-experiencing (RE), avoidance (AV) or arousal (AR), (Foa et al., 1993).

Item number and content	Subscale
1. Hoe vaak heeft u de afgelopen week last gehad van pijnlijke gedachten of beelden over het trauma, terwijl u er niet aan wilde denken? *	RE
2. Hoe vaak heeft u de afgelopen week onprettige dromen of nachtmerries over de traumatische gebeurtenis gehad? *	RE
3. Hoe vaak heeft u de afgelopen week de ervaring gehad dat de traumatische gebeurtenis er weer was, of dat u handelde of zich net zo voelde als toen? *	RE
4. Hoe vaak heeft u de afgelopen week meegemaakt dat u emotioneel overstuur raakte wanneer u aan de traumatische gebeurtenis herinnerd werd? *	RE
5. Hoe vaak heeft u de afgelopen week lichamelijke reacties gehad (bv. in zweet uitbreken of hartkloppingen) als u aan de traumatische gebeurtenis werd herinnerd? *	RE
6. Hoe vaak heeft u de afgelopen week geprobeerd niet aan de traumatische gebeurtenis te denken of geprobeerd niet de gevoelens te voelen die erbij horen? *	AV
7. Hoe vaak heeft u de afgelopen week geprobeerd om activiteiten, plaatsen of dingen te vermijden die u aan de traumatische gebeurtenis herinneren? *	AV
8. In hoeverre heeft u de afgelopen week moeite gehad om belangrijke delen van wat er gebeurd is (tijdens de traumatische gebeurtenis) te herinneren? *	AV
9. Was u de afgelopen week minder geïnteresseerd in dingen die u gewoonlijk belangrijk of leuk vindt (bv. hobby's, sociale activiteiten)? *	AV
10. Voelde u zich de afgelopen week op een afstand of afgesneden van andere mensen? *	AV
11. Voelde u zich de afgelopen week gevoelloos (bv. niet kunnen huilen, niet reageren, onmogelijk om gevoelens van liefde te voelen)? *	AV
12. In hoeverre voelde u de afgelopen week dat uw toekomstplannen of verlangens de grond	AV

in geboord zijn t.g.v. de traumatische gebeurtenis (bv. nooit kunnen werken of carrière maken, geen gelukkige relatie kunnen hebben, geen gelukkige kinderen kunnen hebben, niet lang zullen leven)? *

13. Hoe vaak heeft u de afgelopen week problemen gehad met inslapen of doorslapen? *   AR

14. In hoeverre heeft u de afgelopen week last gehad van snel geïrriteerd zijn of van woedeuitbarstingen? *   AR

15. In hoeverre heeft u de afgelopen week moeilijkheden gehad met concentreren (bv. de draad kwijtraken tijdens een gesprek, de tv niet meer kunnen volgen, niet meer weten wat u zojuist gelezen heeft)? *   AR

16 Was u de afgelopen week erg waakzaam, of op uw hoede (bv. controleren of er niemand in de buurt is, u ongemakkelijk voelen wanneer u geen overzicht heeft)? *   AR

17. Was u de afgelopen week erg schrikachtig / snel geschrokken? *   AR

18. Voelde u zich de afgelopen week erg schuldig over de traumatische gebeurtenis?   -

19. Heeft u zich de afgelopen week erg geschaamd over wat er gebeurd is?   -

20. Had u de afgelopen week veel gevoelens van kwaadheid over wat er gebeurd is?   -

21. Was u m.n. kwaad op uzelf of op anderen? (één antwoord)   -

22. In hoeverre was u er de afgelopen week van overtuigd dat u heel erg gewond zou hebben kunnen raken tijdens de traumatische gebeurtenis?   -

23. In hoeverre was u er de afgelopen week van overtuigd dat uw leven op het spel stond tijdens de traumatische gebeurtenis?   -

Appendix B - Alternative diagnoses given by the clinicians, the extended version of Table 12

Table 12

Extended table. Number of patients diagnosed with alternative primary diagnoses from PTSD (according to the SCID-I) and the correspondence to the SCID-I diagnoses (besides PTSD). When the clinicians did not establish the diagnosis of PTSD, they instead mostly diagnosed a depressive disorder, OCD, anxiety disorder NAO, acute stress disorder, agoraphobia, generalized anxiety disorder or a social anxiety disorder.

Diagnosis by clinician	Number of times diagnosed by the clinician (N)	Consistent with the SCID-I?
No diagnosis	6	Yes: 3 No: 3
Depression	21	Yes: 18 No: 3
Obsessive compulsive disorder (OCD)	6	Yes: 2 No: 4
Anxiety disorder NAO	5	Yes: 0 No: 5
Acute stress disorder	4	Yes: 0 No: 4
Panic disorder without agoraphobia	4	Yes: 3 No:1
Panic disorder with agoraphobia	4	Yes: 4 No: 0
Generalised anxiety disorder (GAS)	4	Yes: 2 No: 2
Social anxiety disorder	4	Yes: 2 No: 2
Specific phobia	3	Yes: 2

		No: 1
Substance dependence/abuse	2	Yes: 2 No: 0
Adjustment disorder	2	Yes: 1 No: 2
Attention deficit hyperactivity disorder (ADHD)	2	Yes: 2 No: 0
Eating disorder	2	Yes: 0 No: 2
Autism	2	Yes: 0 No: 2
Psychotic disorder	1	Yes: 1 No: 0
Somatoform disorder	1	Yes: 1 No: 0
Bipolar disorder	1	Yes: 1 No: 1
Cognitive disorder	1	Yes: 0 No: 1
Excoriation disorder	1	Yes: 0 No: 0