



**Universiteit Utrecht**

Stichting **Centrum '45**



**Identifying profiles of post-treatment functioning among patients treated  
for psychotrauma related complaints**

Clinical Psychology Master's Thesis

by

Rixtia Leenstra

“Essentially, all models are wrong, but some are useful.” (Box, 1987, p. 424)

Student name: Rixtia Leenstra

Solis ID: 5931487

Master Clinical Psychology

Faculty of Social Sciences

Date: April 30<sup>th</sup>, 2017

1st supervisor: prof. dr. P. Boelen, Utrecht University

2nd supervisor: dr. N. van der Aa, Foundation Centre '45

In collaboration with Foundation Centre '45, Diemen, the Netherlands

## **Abstract**

Research in treatment outcomes shows that some patients seeking help following traumatic events do not improve or even become worse. Our understanding of this issue may improve if we include more measures in our outcome evaluation, moving beyond the current predominant focus on symptom measures. This study aimed to identify latent subgroups based on homogeneous outcome patterns across multiple outcome domains. Participants were patients with trauma-related complaints who had received trauma-focused therapy (N=224; 43% with tentative PTSD diagnosis). Treatment outcome was evaluated using the Brief Symptom Inventory scale (BSI), PTSD Checklist-5 (PCL-5), Outcome Questionnaire (OQ-45), Cantril's Anchoring Striving Scale (CLL) and Short Form-36 (RAND-36;SF-36). Latent profile analysis (LPA) was performed. A model with four different profiles of treatment outcomes was proposed: a poor daily functioning profile (30%), an overall positive outcomes profile (30%), an overall negative outcomes profile (33%) and a strong daily functioning profile (7%). With the preliminary results of distinctive profiles in treatment outcomes, the present study emphasizes the heterogeneity of psychological recovery, adds to the enhancement of clinical practice, and provides a springboard for future person-centered studies in treatment outcomes.

## **Acknowledgement**

This thesis is the result of an intensive period of seven months, including many hours work, even more coffees, and an infinite amount of learning. Looking back on it, I must thank two people that were essential to its completion. I would like to express my gratitude to Niels van der Aa, methodologist at the foundation Centre '45 in Diemen, firstly for giving me the opportunity to embark on this project, and also for his statistical expertise, kind help and supervision. The same goes for Paul Boelen, professor at Utrecht University, I thank him for the supervision meetings, critical feedback, pop-quiz type questions and encouragement. I want to finish by saying I very much enjoyed being a research intern at the Centre and getting to meet inspiring researchers, clinicians and other interns.

## Table of contents

Introduction .....	6
Method .....	9
Procedure and participants .....	9
Measures.....	10
Data analysis.....	11
Results .....	13
Model selection .....	13
Outcome profiles .....	14
Study overview.....	21
Limitations .....	21
Implications and future research .....	22
Conclusion.....	23
References .....	24

## **Introduction**

Despite the fact that many types of psychotherapy are empirically supported, research shows that a small group of patients does not respond favorably to treatment and even 5 to 10 percent of patients worsen (Slade, Lambert, Harmon Smart & Bailey, 2008). This problem is especially characteristic of Post-Traumatic Stress Disorder (PTSD), where patients often experience residual symptoms after treatment, and in rare cases will suffer chronically (Bradley, Greene, Russ, Dutra, & Westen, 2005; King et al, 2014). Researchers agree that there is a need to improve our understanding of how, when and for whom treatments have a meaningful effect (King & Resick, 2014). Specifically, in what ways a patient may benefit from treatment. Some patients for instance, could be relieved from symptoms targeted in treatment (e.g., re-experiencing in PTSD) while still suffering from general aspects of dysfunction, whereas others may improve significantly regarding satisfaction with life. Although variation in outcomes is common, most therapies were designed to specifically address symptoms, and evaluations of outcomes or effectiveness of treatments is predominantly focused on symptom reduction. This paints a one-dimensional picture: that a disorder equals the presence of a certain set of symptoms. Some studies, however, suggest that assessing symptoms alone would not be enough to form a holistic (i.e., complete) image of a patient's status or progress. Next to symptoms, interpersonal or emotional regulation problems can negatively impact a person's well-being by causing dysfunction at work, in marriage or parenting, a low perception of social support, or high reactivity (Cloitre et al., 2010). In recent years, the heterogeneity of trauma-related disorders and treatment outcomes has gained attention. For example, a study found that short- and long-term responses to potentially traumatic events are much more varied than believed and that the factors contributing to healthy functioning in the aftermath are heterogeneous in nature (Bonanno & Mancini, 2012). But, it appears that attempts to disentangle this subject display the traditional focus on symptom presentations of PTSD (Dalenberg, Glaser, & Alhassoon, 2012). Examples of this are studies in the "complex PTSD" category, "a dissociative subtype of PTSD" or classifications of symptom severity (Armour et al., 2015; Cloitre et al., 2013). The attention for heterogeneity is promising, but a multidimensional view of treatment outcomes is lacking.

It has been proposed that clinicians should assess how a patient functions relative to how a 'healthy' person would function, and how this changes during treatment (Lambert, 2004; Wells, Burlingame, Lambert, Hoag, & Hope, 1996). Functioning levels such as interpersonal, emotional, or physical functioning, are important aspects of wellbeing that are affected by mental illness (Armour et al., 2015) as some studies make clear. Firstly, there is an association between PTSD and increased odds of bed days, poorer physical health, and unemployment, found even after adjusting for medical and psychiatric comorbidities (Zatzick et al., 1997) there is also an association between number of lifetime traumas and number of PTSD symptoms with poorer physical, role, and cognitive functioning (Leserman et al., 2005) and lastly, Cloitre et al. (2005) of the aforementioned study recommended to

include factors such as emotion regulation and interpersonal problems when making treatment and rehabilitation plans.

Another outcome domain that mental illness impacts upon is quality of life (QOL). Quality of life is defined as an overall assessment of life according to personally chosen criteria and depending on a patient's ability to have a life that fulfils personal needs (Shin & Johnson, 1978). Anxiety disorders are generally associated with reduction in quality of life, and this is especially true for PTSD, a meta-analysis reports (Charney & Marx, 2012). It has been hypothesized that, given the complex relationship between PTSD and physical health, the effects of PTSD symptom improvement on physical health might be delayed because initial symptom change was unrelated to later changes in quality of life. Based on this result it was suggested that treatments designed to enhance quality of life could have additional benefit to patients beyond the benefit resulting from PTSD treatment. This pointed to a need for a measure of QOL in investigations of trauma therapy (Schnurr, Hayes, Lunney, McFall, & Uddo, 2006).

The recovery of a patient can be seen as a trajectory in which normal functioning temporarily gives way to threshold or subthreshold psychopathology before returning to a regular level. A description as such implies that symptoms need to diminish in order for the person to function normally again. However, there is no evidence to date explaining how this relationship between symptoms and problems like dysfunctioning works. In one study, researchers looked at the relative contribution of problems in emotion regulation and interpersonal functioning compared to PTSD symptoms in predicting functional impairment among women with childhood abuse histories. The study reported that severity of PTSD symptoms was a significant predictor of functional impairment. But, symptom level was not the only predictor. In addition, emotion regulation and interpersonal problems were both significant predictors and contributors to functional impairment equal to that of PTSD symptoms (Cloitre, Miranda, Stovall-McClough, & Han, 2005)

The abovementioned studies illustrate that mental health problems after trauma vary widely in the severity and also in the nature (e.g., symptoms, functioning, QOL) of complaints. Additionally, the relationship among complaints in different domains cannot be assumed, and research continues to focus on symptoms while there is sufficient evidence for the multidimensionality of trauma-related psychopathology. Heterogeneity is often investigated by looking at the relationships among people (i.e., 'person-centered research') instead of variables. This study will analyze posttreatment scores of patients treated for moderate to severe trauma-reactions in a person-centered manner, and thereby include multiple outcome domains. Other than the idea that some patients end their course of treatment with less complaints than others, it is not known to date what outcomes look like. Latent profile analysis (LPA), a multivariate statistical technique, will enable us to identify subgroups of individuals based on similarities in their responses to a set of continuous observable indicators (Vermunt & Magidson, 2002). Our primary aim is investigating the relevance of including multiple wellbeing-domains such as role functioning, daily functioning and quality of life in evaluations of

post-treatment scores. The first research question then is: *is it valuable to include different domains when analyzing patient's posttreatment scores?* Here, 'valuable' indicates the extent to which this differentiates the post-treatment picture. For instance, if posttreatment scores appear similar across different domains (e.g., only low scores, only high scores) then it is not useful to include more domains and therefore not valuable. Our secondary aim is possibly revealing the existence of latent subgroups based on patterns in multifaceted posttreatment scores and provide a more differentiated image of evaluation of outcomes compared to the traditional symptom-focused view. Hence the second question: *can homogeneous subgroups based on posttreatment scores be identified?* This study is cross-sectional and uses data acquired through Routine Outcome Monitoring (ROM; Ellwood, 1988). Profiles of treatment outcomes will be defined by post-treatment scores on seven different measures (i.e., the indicators) on three different outcome domains. First, general psychopathology and PTSD symptoms in the domain of symptoms and complaints, second, the functioning domain with measures of societal and interpersonal functioning, and physical and emotional role functioning and finally the domain of quality of life. Findings from this study can ultimately contribute to theory and clinically applicable information with the aim to improve trauma related treatment.

## Method

### Procedure and participants

Data for these analyses were obtained from measures administered as a part of the Routine Outcome Monitoring (ROM) procedure at foundation Centrum '45 between 2001 and 2016. Foundation Centrum '45 is a specialized psycho-trauma treatment centre in the Netherlands. Patients admitted there will have experienced one or more life events that carried high risk of traumatization, will have needed specialized trauma-focused psychotherapy, sometimes in combination with pharmacotherapy. Table 1 shows the background of the participating patients. The ROM procedure is that patients fill out questionnaires at intake, after every six months and when treatment is ended. In the ROM database, pre- and post-treatment data were available for only 38 participants. Because we found a sample size of  $N = 38$  to be too small, we chose to include post-treatment scores only, meaning patients who had completed treatment. 'Completed treatment' does not indicate treatment termination by the patient, but a discharge of the patient by the clinician, when it was believed that continuation would not lead to any further improvement. This is regardless of whether patient or clinician deem the treatment successful. Average duration of treatment in months was  $M=34.12$  ( $SD=27.81$ ). The total sample included 224 patients (59.8% male, 40.2% female). Mean ( $M$ ) age of participants was 48.50 ( $SD = 11.36$ ); the country of origin, for the majority of participants, was the Netherlands (74.6%). As seen in Table 1, the majority of patients in this study came from a background categorized as: 'post-war generation' (their parents had experienced WWII), 'veterans' and 'profession-related' (one or more traumatic experiences through a job such as police officer). Others were refugees, asylum seekers or non-documented people. There were 96 (42.9%) patients with a provisional diagnosis of PTSD. Mastery of the Dutch language was an inclusion criterion because Dutch questionnaires were used.

Table 1

*Patient's background*

Background	N	(%)
Total sample	224	(100)
Post-war generation	75	(33)
Veterans	51	(23)
Profession-related traumatized individuals (e.g., police)	39	(17)
Refugees, asylum seekers and non-documented individuals	24	(11)
Other	35	(16)

## Measures

**Brief Symptom Inventory (BSI; Derogatis, 1993; De Beurs, 2011)** The BSI is a reliable and valid self-report measure of general psychopathology and is used here as an indicator variable for the symptoms domain. It can be used to assess the nature and severity of the complaints, measuring 9 primary symptom dimensions of psychopathology (e.g., depression “*feeling no interest in things*”; hostility “*having urges to break or smash things*”). The total of 53 items was rated on a 5-point scale, ranging from 0 (*not at all*) to 4 (*extremely*) with higher scores reflecting greater severity of symptoms. Only the total score, which equals the average score on all items, was used. In this sample the reliability reflected by Cronbach’s alpha was .98 and we used the Dutch version of the BSI (De Beurs, 2011).

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

The self-report measure PCL-5 was used to indicate outcomes in the symptoms domain and measured PTSD-symptoms according to the 20 PTSD symptoms of the DSM-5 (e.g., “*In the past month, how much were you bothered by: "Repeated, disturbing, and unwanted memories of the stressful experience?"*”). The 20 items were rated on a 5-point Likert scale (“*Not at all,*” “*A little bit,*” “*Moderately,*” “*Quite a bit,*” and “*Extremely*”). The DSM-5 symptom clusters for PTSD (*intrusion, avoidance, negative alterations in cognitions and mood, and alterations in arousal and reactivity*) are represented by subscales. The sum of the 20 item scores reflects the total symptom severity score (0–80) which was used as the indicator variable. To our knowledge, information about the psychometric qualities of the PCL-5 specifically has not yet been published. However, the original PCL has adequate reliability ( $\alpha = .94$ ; test–retest  $r = .88$ ) in various trauma-exposed populations (Ruggiero, Del Ben, Scotti, & Rabalais, 2003). A Cronbach’s alpha of .97 was found for this sample.

### **Outcome Questionnaire-45 (OQ45; Lambert, Gregersen & Burlingame, 2004; Wells, Burlingame, Lambert, Hoag & Hope, 1996)**

The Outcome Questionnaire-45 (OQ-45) was originally designed to measure client progress in therapy. It assesses three aspects of the client’s life: subjective discomfort or symptoms; problems in interpersonal relationships and problems in social role performance. Items are scored on a 5-point scale and the total score yields a range of possible scores of 0–180, where higher values indicate higher levels of client distress and pathology and lower scores indicate adequate or good therapeutic progress. Two subscales of the OQ-45 were utilized to measure the domain of role functioning: the Interpersonal Relationships scale (IR), as indicator for interpersonal functioning, and the Social Role scale (SR) as indicator for societal functioning. *IR*, consisting of 11 items, reflected interpersonal dysfunction and dissatisfaction with the quality of relationships. It also measured to which degree loneliness is experienced, or conflicts in the close social circle, since these add to possible dissatisfaction and indicate interpersonal dysfunction. The second subscale *SR*, consisting of 9 items, reflected success and satisfaction with performing a social role (e.g., student, employee, friend).

Therefore it also measured to which degree conflicts or distress existed at work or in education. The OQ-45 has adequate internal consistency ( $r = .93$ ) and three-week test–retest reliability ( $r = 0.84$ ). Concurrent validity is moderate to high ( $r$  values of  $.50 - .85$ ; Boswell, White, Sims, Harrist & Romans, 2013) when correlated with common psychotherapy outcome-measures. The OQ-45 scores are sensitive to change in clients over short time periods, while remaining stable in untreated individuals (Lambert, 2015). The IR and the SR in this sample produced an alpha of  $\alpha = .87$ .

### **RAND 36-Item Health Survey (VanderZee, Sanderman, Heyink & deHaes, 1996)**

The RAND-36 is a self-report questionnaire which contains 36 questions measuring both positive and negative health states, where higher values indicate a better state of health. In this study, two out of the eight subscales were used as indicators for the functioning domain. Firstly, the subscale ‘role limitations due to physical problems’, including 4 items measuring problems with work or other daily activities as a result of physical health problems during the last four weeks (e.g., “*You were limited in the type of work or activity*”; “*You have achieved less than you wanted to*”). Secondly, the subscale ‘role limitations due to emotional problems’: 3 items regarding role limitations due to emotional problems (e.g., “*You have done your work or activities less carefully than you are used to*”). Participants are given the options “yes” or “no” to answer. On both scales, high scores indicate that patients do not feel limited in their daily activities due to problems with a physical or emotional nature specifically. Test-retest reliability for the limitations due to physical problems scale was found to be  $.60$  (after 2 months) and  $.47$  (after 6 months). For the emotional problems scale these were respectively  $.67$  and  $.40$  (VanderZee & Sanderman, 2012). Cronbach’s alpha in this study was  $.92$

### **Cantril’s Ladder of Life (CLL; Cantril, 1965)**

This short self-report questionnaire was used to reflect the quality of life domain. It instructs respondents to rate the quality of their life at present (item 1) and in five years from now (item 2). For this, a scale depicted as a ladder is given with steps ranging from 0 (worst possible life) to 10 (best possible life). Only scores on the first item were used in this study to indicate QOL. Scores 0-4 are seen as weak and reflecting wellbeing that is at risk, scores 5 or 6 are both moderate, and 7 or higher reflect the positive end of the continuum, and “wellbeing that is strong” (Gallup, 2009). Construct validity has been demonstrated in a past study in life satisfaction (Brown, Rawlinson & Hilles, 1981), and the CLL showed moderate to high correlations with other wellbeing measures like the Satisfaction With Life Scale and the Subjective Happiness Scale (Bartels & Boomsma, 2009).

### **Data analysis**

Latent Profile Analysis (LPA) was performed using Mplus version 7.3 (Muthén & Muthén, 2014) on seven continuous indicators (see Table 2). Based on the given indicators, LPA can calculate if homogeneous subgroups exist in a seemingly heterogeneous group of people. To facilitate

interpretation of scores from different measurements, all scores were standardized to a 0-100 scale. First, several models with increasing number of class-solutions were fit to the data. The robust maximum likelihood estimator (MLR) with full information maximum likelihood estimation was used to include participants with missing data. To avoid local likelihood maxima, we requested 50 initial stage iterations next to random sets of 500 starting values in the first, and 50 in the second step of optimization. Second, the model fit indices provided through likelihood ratio tests were used to determine the optimal number of latent classes. The Bootstrap Likelihood Ratio Test (BLRT) and the Vuong-Lo-Mendell-Rubin Test (LMR LRT), compare the fit of the estimated model with a model with one class less. A significant result suggests that the model with  $N$  profiles is better than the one with one class less ( $N-1$  profiles). Nonsignificant results suggest that a more parsimonious model than the one extracted is preferred. For 500 bootstrap samples 50 sets of starting values were requested for the first step and 20 sets for the second step of optimization, again to avoid local likelihood maxima. For the model-selection process, five criteria were considered—the BLRT, the LMR LRT, Bayesian Information Criterion (BIC), entropy, and how interpretable the profiles are theoretically (Wang & Wang, 2012). With regard to BIC, the model with a lower value is preferred. Entropy indicates the certainty with which the subjects can be classified into the classes. Entropy above .6 is generally agreed on as medium and above .8 as high (Clark, 2010). In case of inconsistency between the model fit indices, BLRT and BIC are recommended to be the best indicators (Nylund, Asparouhov, & Muthen, 2007). Latent class analyses attempt to find a meaningful model, that is clearly interpretable, showing distinctive outcome patterns for classes with reasonable proportions. Therefore, interpretation weighed heavily in the selection process.

Table 2  
*Overview of indicator variables*

Outcome domain	Measure name	Parts included in analysis
Psychological complaints and PTSD symptoms	BSI and PCL-5	Total scores (1&2)
Societal functioning	OQ-45	(3) Social Role and (4) Interpersonal Relationships
Role functioning	RAND-36	Role limitations due to (5) physical problems and (6) emotional problems
Quality of life	CLL	(7) Rating life presently (item 1)

*Note. High values of OQ-45 reflect societal dysfunction. High values are positive only for RAND-36 and CLL scores, reflecting respectively an absence of limitations, and a positive evaluation of life.*

## Results

### Model selection

In total, six class models were fitted to the data. The LMR LRT suggested that adding a seventh class did not improve model fit ( $p = .4512$ ). Table 3 gives the fit indices for the 1- to 6-class model. As can be seen in Table 3, all six models had excellent entropy values (above .8). The bootstrapped likelihood ratio (BLR) test produced significant values at the  $p < .01$  level for all six models. The LMR LRT, that compares models with each other to assess what number of classes is correct, gave significant results at  $p < .05$  for the 2-, 3-, and 6-class models. This indicates that the 2-class model fit was better than the 1-class solution, the 3-class model fit better than the 2-, and the 6-class model fit better than the 5-class solution. The similar BLR test gave significant results for all six models. The 6-class model additionally had the lowest BIC value ( $= 13302.491$ ). However, looking at the outcome patterns for the 5-class and the 6-class models in figure 1 it was seen that these models included some relatively smaller classes ( $< 10\%$  of the sample) with indistinguishable patterns. Namely, in the 5-class model, class 5 was only marginally different from class 2, as was the case for classes 6 and 4 in the 6-class model. Therefore, the 6-, or 5-class solutions were not preferred. The 2-class model showed insufficient goodness-of-fit results. Although it had a significant LMR LRT result it also showed a relatively high BIC. In comparison, the 3-class model produced a visible decline in BIC value while the LMR LRT was also significant for  $p < .05$ . Adding to this, the two subsequent models (3- and 4-class) identified outcome patterns distinctive from the ones shown by the 2-class model which suggested that 2 classes do not convey the most possible information. With figures 1B and 1C, the interpretability of the classes from the 3- and 4-class models were compared. Firstly, for the 3-class model: class 1 showed a pattern of good outcomes, with low pathology, low symptoms and low dysfunctioning in combination with an absence of physical or emotional problems causing role limitation and an average QOL. Class 2 showed a pattern with outcomes exactly opposite to the class 1-pattern, considering the relatively high scores on symptoms and dysfunctioning and relatively low scores on daily functioning and quality of life; class 3 showed moderate scores on psychopathology, symptoms and dysfunctioning in combination with feeling limited in role functioning but a reasonable score on QOL. Secondly, the 4-class solution added a fourth class of patients with moderate symptoms and dysfunctioning, but who did not feel limited in role functioning and score moderately on QOL. This pattern is visibly different from class 3 in the 3-class model, which strongly suggested that the 4-class model added a meaningful class with class 4 (See figure 1C). In the 4-class model daily functioning appeared to be a differentiating factor, showing low scores for class 2 and high scores for class 4 in this model (figure 1C). Figure 2 showed results compared to norm-scores in the regular and clinical population. Figures 2E and 2F confirmed that, for class 4, daily functioning scores were relatively positive (“not feeling limited”), and that quality of life (figure 2G) was positive for most people with 13% scoring low, 67% scoring moderately and 20% scoring high. Therefore, based

on figure 1 and figure 2 it was concluded that the 4<sup>th</sup> class had added value in terms of theoretical interpretation and the 4-class solution was the best fitting model.

### Outcome profiles

The first profile, profile 1 (P1, 30%), included relatively low scores on interpersonal and functioning subscales which means that these people do not experience positive relationships or have trouble with them and additionally feel limited in functioning (e.g., in work). Contrastingly this group did not show quality of life scores in the lower range but the high and moderate range (57% of people in P1). Less than one-third have a provisional PTSD diagnosis. So, the strengths (experiencing symptoms and satisfaction with life) and the weaknesses (dysfunctioning in everyday tasks and in a societal role) of this group were clearly observable. Profile 2 (P2, 30%) showed low scores in interpersonal functioning. That negative outcome aside, there was little psychopathology, the least provisional PTSD diagnoses, high quality of life, and moderate societal dysfunctioning. Therefore, this is not a group with a particular strength or weakness, but showing positive outcomes across domains. The majority (60% of people in P2) scored high or very high, however, in interpersonal dysfunctioning. In profile 3 (P3, 33%) no members scored 'very high' or even 'high' for interpersonal dysfunctioning. In fact, they had the highest percentage of people in the low to very-low range. On the other hand, the large majority of people in P3 scored negatively for all other indicators. Opposite to what can be seen in P2, this relatively negative outcome profile is defined by people with strong social bonds and role functioning but a pressing need for improvement in other areas. Lastly, profile 4 (P4, 7%) appeared average in outcomes. Yet, these moderate symptoms and complaints (psychopathology, PTSD symptoms and role functioning) opposed the relatively good scores for perception of daily functioning (one-third or less of people reported feeling limited) and a moderate score for quality of life (67% of people scored in the moderate range). In P2, scores in daily functioning and QOL are similarly high but were a contrast to symptoms and psychopathology being largely absent.

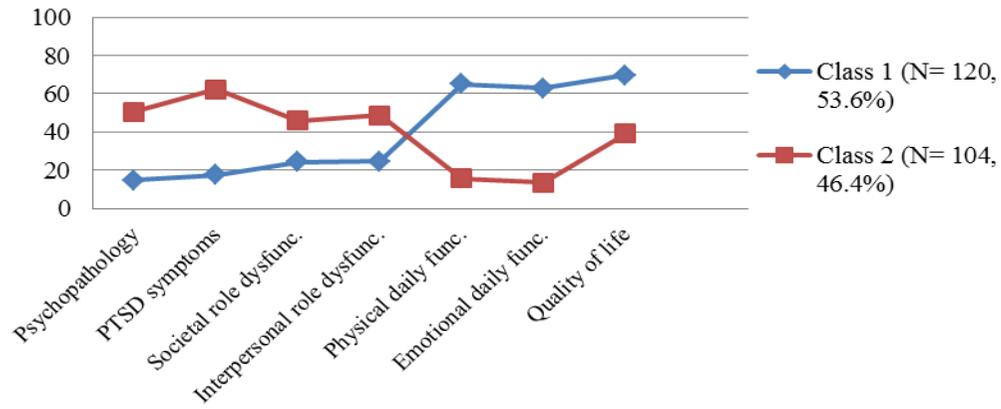
Table 3

*Model fit indices for each class model (N = 224)*

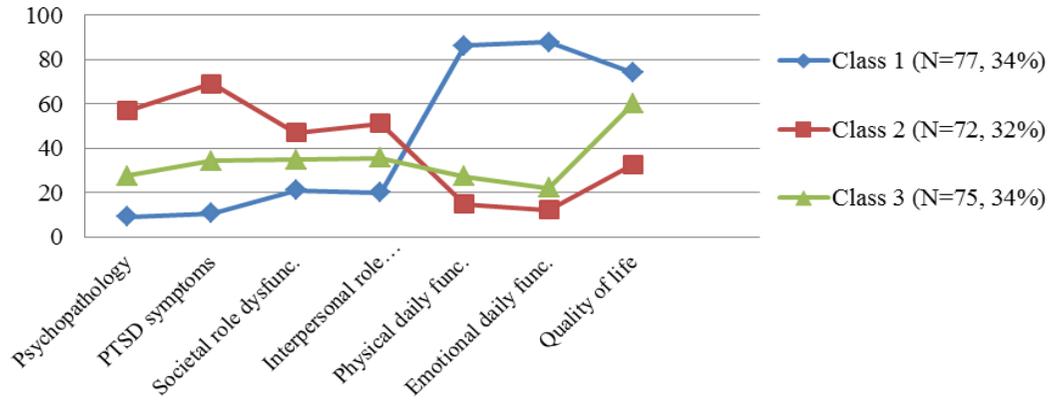
Model	BIC	Entropy	Log-likelihood	BLRT		LMR LRT	
				-2LL difference	p-value	Value	p-value
1 class	14626.024	1.000	-7.275.130	--	--	--	--
2 classes	13804.431	.941	-6.842.687	864.886	$p < .01$	845.360	$p < .01$
3 classes	13515.935	.931	-6.676.793	331.789	$p < .01$	324.299	.0140
<b>4 classes</b>	<b>13405.539</b>	<b>.957</b>	<b>-6.599.948</b>	<b>153.689</b>	<b><math>p &lt; .01</math></b>	<b>153.689</b>	<b>.0795</b>
5 classes	13328.177	.937	-6.539.621	120.655	$p < .01$	117.931	.0759
6 classes	13302.491	.934	-6.505.131	68.979	$p < .01$	67.422	$p = .0327$

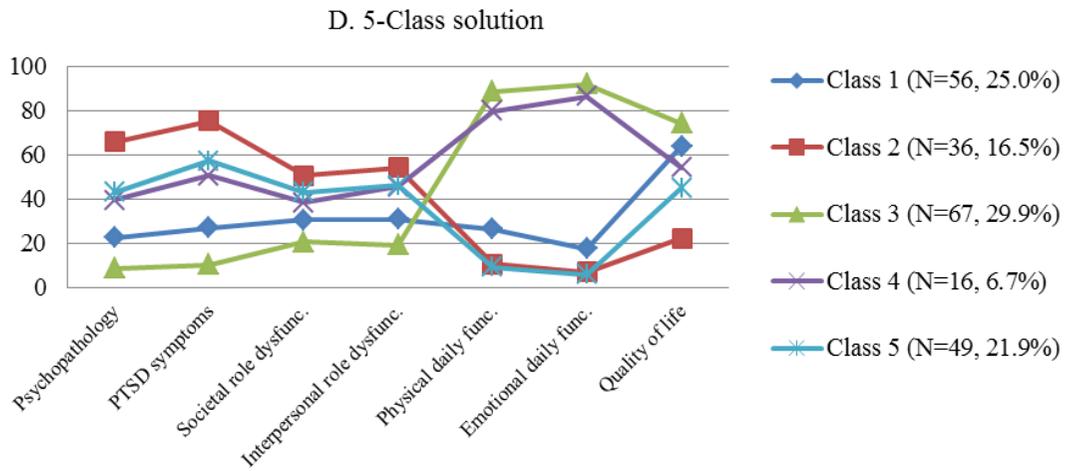
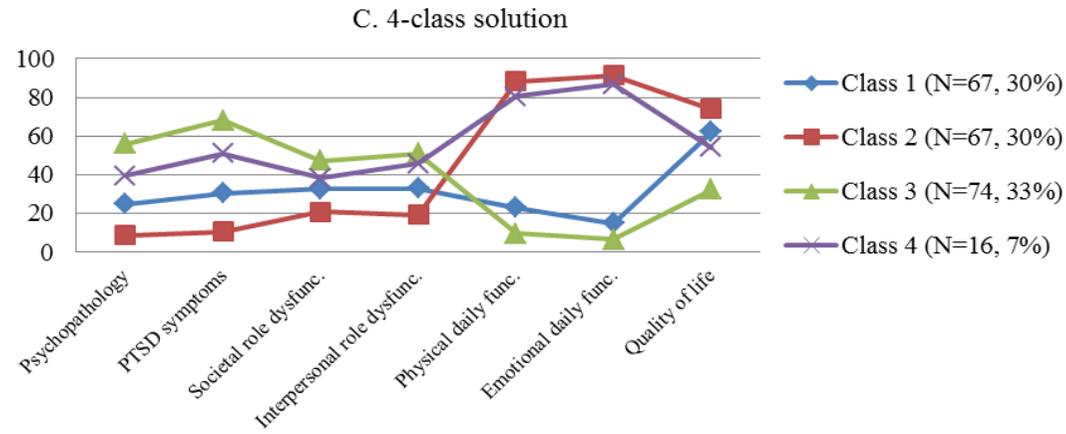
*Note.* BIC = Bayesian Information Criterion; BLRT = Parametric bootstrapped likelihood ratio test; -2LL difference = 2 times log-likelihood difference between a N class solution and N – 1 class solution; LMR LRT = Lo-Mendell-Rubin adjusted likelihood ratio test. Best fitting model is printed in bold.

A. 2-class solution



B. 3-class solution





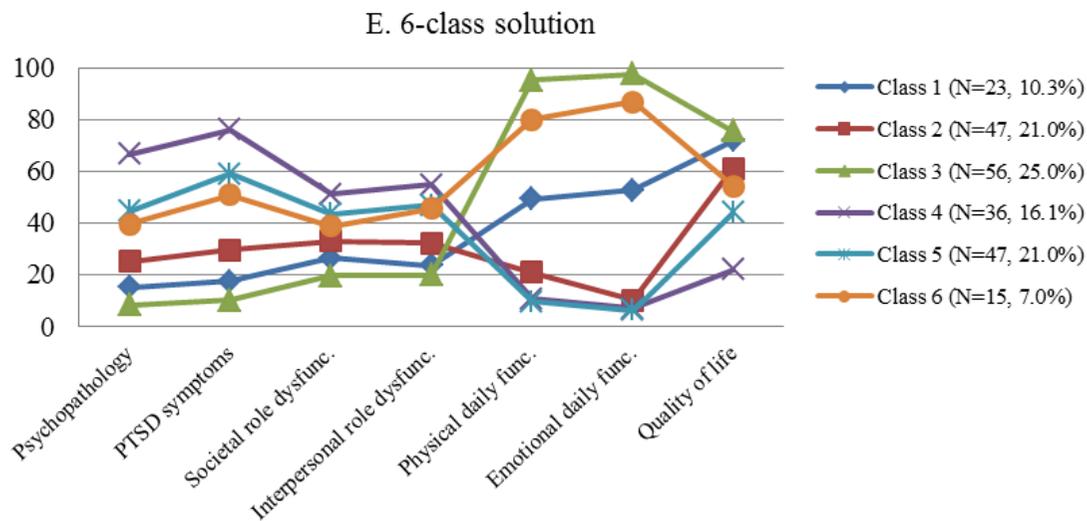
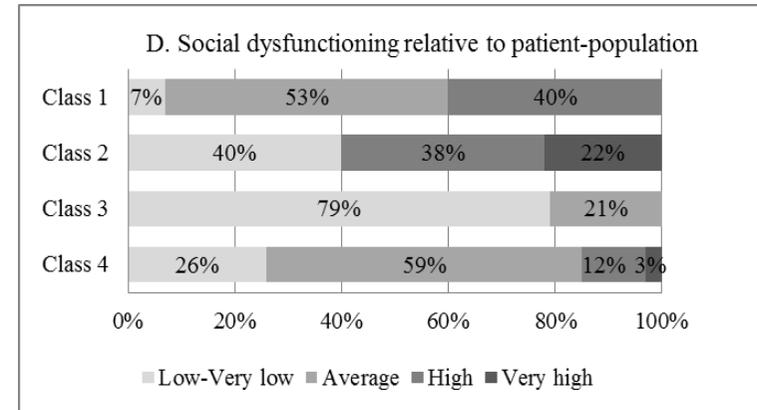
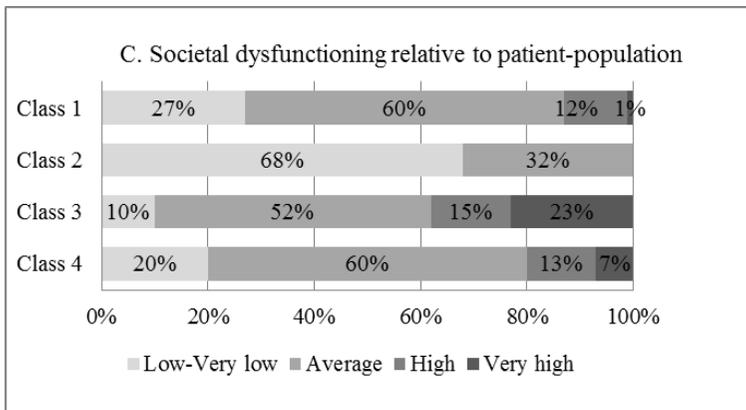
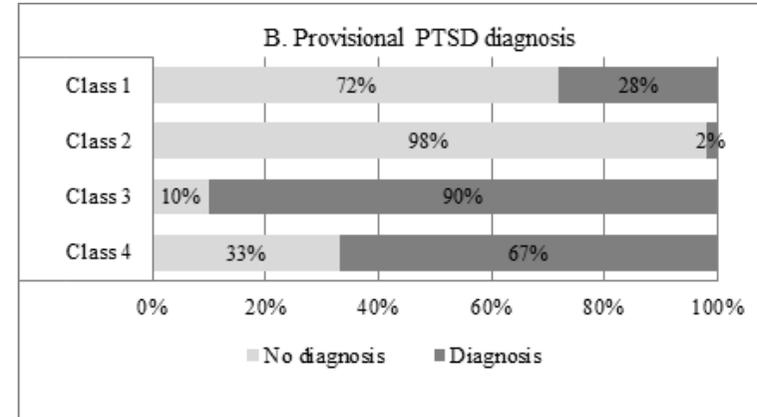
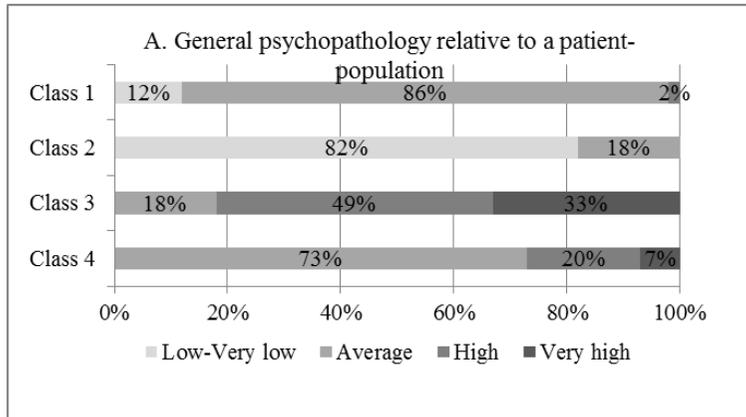


Figure 1. Mean scores on the 7 indicator variables for the 2-, 3-, 4-, 5- and 6-class solution. Note. Psychopathology=BSI; PTSD symptoms=PCL-5; Societal role dysfunctioning= OQ-45 societal subscale; Interpersonal role dysfunctioning=OQ-45 interpersonal relationships subscale; Physical daily functioning = RAND-36 feeling limited due to physical problems subscale; Emotional daily functioning=RAND-36 not feeling limited due to emotional problems; Quality of life=CLL satisfaction with life presently subscale.



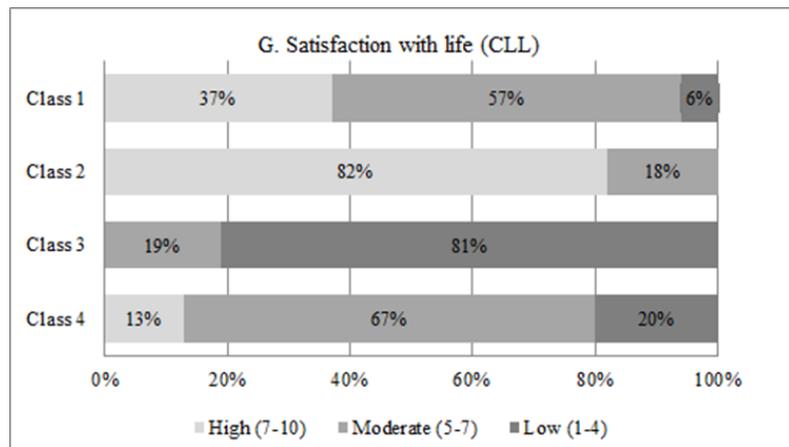
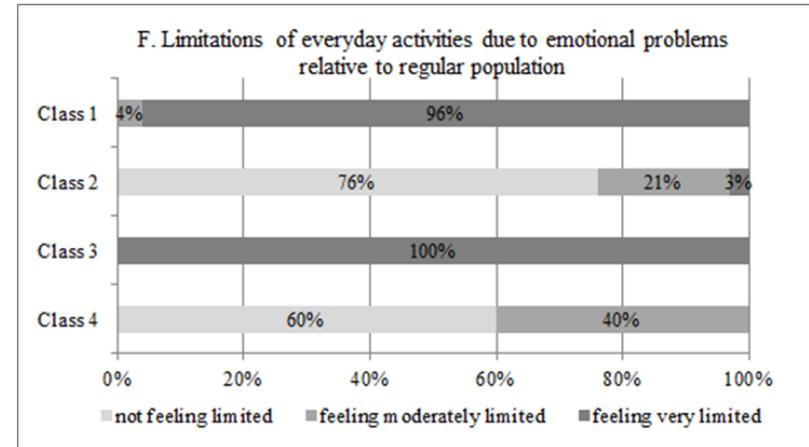
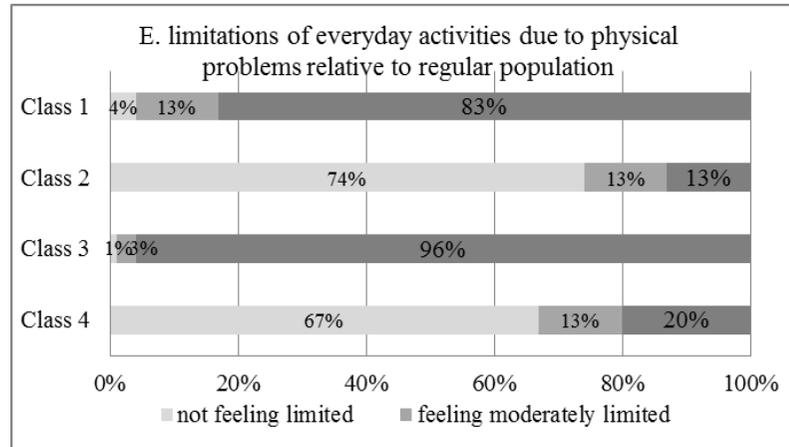


Figure 2. Class average scores relative to norm-scores for each measure. Note. A=BSI scale; B=PCL-5 scale; C=Societal role functioning subscale OQ-45; D=Interpersonal relationships subscale OQ-45; E=limitations due to physical problems subscale RAND-36; F= limitation due to emotional problems subscale RAND-36; G= satisfaction with life presently subscale CLL.

## Discussion

### Study overview

This was the first LPA study done in post-treatment scores, identifying homogeneous profiles in a group of patients with psychotrauma related complaints. The 4-profile model was the best fit in this study. The four profiles are defined as: *'poor daily functioning'* (P1), *'overall positive outcomes'* (P2), *'overall negative outcomes'* (P3) and *'strong daily functioning'* (P4). The 4-class model is superior to other class models regarding theoretical interpretation. In addition to a positive outcome and a negative outcome profile, this model adds two meaningful profiles because the daily functioning scale visibly differentiates between them. From a purely statistical viewpoint, the 4-class model is not superior to other models, as the LMR LRT did point towards a three or six-class solution. However, the optimal number of classes was based on consideration of not just statistical results but all required criteria: the research question, parsimony, theory, and the substantive meaning of each solution (Berlin, Williams, & Parra, 2013). What was observed across all six class-solutions was the presence of at least one profile with overall positive outcomes and one profile with overall negative outcomes, and that overall scores of the OQ-45 showed little variability between classes (See figure 1 in the Results section). The line graphs delineate that outcomes differ across and between domains for groups of patients. These results weaken the notion that when symptoms are reduced, the patient is without complaints, which has been described in the literature (Bradley et al., 2005). Profile 2 shows how symptom-level can be low while societal- and interpersonal dysfunctioning levels can be high (see figure 1 and 2 in result section). The first research question is answered affirmatively: there is value in analyzing outcomes including different outcome domains. We compared our findings to norm-scores of male and female mental health care outpatients and mentally healthy people in the Netherlands which enabled us to place the findings in perspective and label the profiles. The second research question regarding the existence of subgroups is therefore confirmed. These relative scores of these four profiles seen in figure 2 suggest that it may be wrong to assume that when symptom scores are high post-treatment, outcomes in more subjective domains like QOL will be low. Especially profile 4 contests this notion, showing a subgroup of people ( $N= 16$ ) with an ability to function despite mental health symptoms.

### Limitations

There are several limitations to be named. Although this analysis of outcomes provides new insights, cross-sectional studies do not allow monitoring a trend over time which is a weakness of the design. Because of the design and the use of posttreatment scores, we cannot shed light on patient's recovery trajectories or infer any causations. Additionally, the choice of indicator variables was subject to what measures had been included in the ROM battery. The included measures are self-report measures, so the degree to which people have answered truthfully may also influence outcomes. Further, possibly because of a high ( $\geq .89$ ) class separation in this sample the p-value for the BLRT

turned out significant for every model. BLRT can generally detect the true model when the separation between classes is moderate to high but not when the class separation is low (Tekle, Gudicha, & Vermunt, 2016). It is possible that the BLRT lacked accuracy, and therefore barely influenced the model selection in this study. Finally, in terms of external validity it should be noted that the sample concerned treatment-seeking and Dutch-speaking individuals. The majority were Dutch nationals, of the post-war generation, or with profession related trauma

### Implications and future research

The identification of latent subgroups help to understand heterogeneity in treatment outcomes and recovery in the field of psychotrauma, where focus seems to be mainly on PTSD and its subtypes or symptom presentations (Dalenberg et al., 2012). There is an abiding issue in clinical psychology that patients are sometimes not responding to treatment in the intended manner (Slade et al., 2008). PTSD patients specifically have chances of experiencing residual symptoms. Even when criteria for PTSD are no longer met, residual symptoms are a problem, as studies show that subthreshold PTSD is associated with significant impairment in work and social functioning as well as suicide attempts (Bradley et al., 2005). Slade et al. (2008) proposed that formal tracking of patient progress is one way to facilitate the data that helps alert clinicians of unintended treatment effects and improve outcomes. But another part of the problem is that treatment outcomes are ill-defined. There appears to be a need for such broader perspectives on recovery in mental health. Defining of the criteria for treatment response remains an essential task (Macher & Crocq, 2004). Commonly treatment outcomes are categorized as either responding, not responding, or worsening, in terms of symptom reduction. Dalenberg et al. (2012) propose that investigating subtypes of PTSD could enhance treatment modification and effectiveness. What we did then, was apply this to treatment outcomes. We were able to identify profiles within a set of post-treatment scores. Although are not able to speak of ‘treatment effects’, a strong feature of this study is the recommendation of a broader perspective in evaluating outcomes. This means including domains such as role- and daily functioning, and quality of life, and serves as an important take-away message for clinicians and policy-makers in healthcare. If such domains are not included in an evaluation of outcomes or effects, the result may be an incomplete picture of the person’s well-being. Broader perspectives on well-being and recovery are advocated in emerging conceptual frameworks, like ‘CHIME’ (Connectedness, Hope, Identity, Meaning in life and Empowerment) which describes recovery processes and stages using different characteristics (Leamy, Bird, le Boutillier, Williams & Slade, 2011). By synthesizing people’s experiences in recovery CHIME attempts to provide an empirical basis for recovery-oriented research. This is contrasting with categorizing patient’s responses or outcomes in terms of symptom reduction. Finally, understanding what factors underlie negative treatment outcomes is a necessary part of the solution. Future research can build on our findings by investigating which factors predict membership of an outcome profile, why the domain of functioning is a differentiating factor, identify other mean differences in outcomes

across latent profiles, or attempt to describe the extent to which latent class membership moderates the relationship between two or more variables.

## Conclusion

Studying the severity but also the nature of complaints across multiple domains may hold the key to enhancing treatment outcomes. The findings of this study emphasize the relevancy of looking at differences in outcomes across symptomatic, functioning, role performance and quality of life domains. These are preliminary findings but can be a starting point for future research in developments of trauma-focused psychotherapy. It has been recommended that treatment development and selection of best-fitting treatments take into account the diversity of symptom presentations (Elhai Naifeh, Forbes, Ractliffe, & Tamburrino, 2011). To this statement, we would add: the diversity of presentations of dysfunctioning and quality of life scores.

## References

- Armour, C., Contractor, A., Elhai, J.D., Stringer, M., Lyle, G., Forbes D., & Richardson, J.D. (2015). Identifying latent profiles of posttraumatic stress and major depression symptoms in Canadian veterans: exploring differences across profiles in health related functioning. *Psychiatry Research*, 228, 1-7.
- Berlin, K. S., Williams, N. A., & Parra, G. R. (2014). An introduction to latent variable mixture modeling (part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of Pediatric Psychology*, 39(2), 174-187.
- DeBeurs, E. (2008). *Brief symptom inventory handleiding*. Leiden: The Netherlands: PITS B.V.
- Bonanno, G. A., & Mancini, A. D. (2012). Beyond resilience and PTSD: Mapping the heterogeneity of responses to potential trauma. *Psychological Trauma: Theory, Research, Practice, and Policy*, 4(1), 74.
- Box, G.E.P. & Draper, N.R. (1987). *Empirical Model-Building and Response Surfaces*. Wiley, p.424.
- Bradley, R., Greene, J., Russ, E., Dutra, L., & Westen, D. (2005). A multidimensional meta-analysis of psychotherapy for PTSD. *American journal of Psychiatry*, 162(2), 214-227.
- Cantril, H. (1965). *Pattern of human concerns*. New Brunswick, NJ: Rutgers University Press.
- Charney, M. E., & Marx, B. P. (2012). Promoting a positive quality of life posttrauma.
- Cloitre, M., Stovall-McClough, K. C., Nooner, K., Zorbas, P., Cherry, S., Jackson, C. L., ... & Petkova, E. (2010). Treatment for PTSD related to childhood abuse: A randomized controlled trial. *American journal of psychiatry*, 167(8), 915-924.
- Cloitre, M., Garvert, D. W., Brewin, C. R., Bryant, R. A., & Maercker, A. (2013). Evidence for proposed ICD-11 PTSD and complex PTSD: A latent profile analysis. *European journal of psychotraumatology*, 4.
- Dalenberg, C. J., Glaser, D., & Alhassoon, O. M. (2012). Statistical support for subtypes in posttraumatic stress disorder: The how and why of subtype analysis. *Depression and Anxiety*, 29(8), 671-678.
- Derogatis, L. R., & Spencer, P. M. (1993). *Brief symptom inventory: BSI*. Upper Saddle River, NJ: Pearson.
- Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, 49(1), 71-75.
- Elhai, J. D., Naifeh, J. A., Forbes, D., Ractliffe, K. C., & Tamburrino, M. (2011). Heterogeneity in clinical presentations of posttraumatic stress disorder among medical patients: testing factor structure variation using factor mixture modeling. *Journal of traumatic stress*, 24(4), 435-443.
- Gallup, G. (2009). *World Poll Methodology*. Technical Report. Washington, DC.
- King, M. W., & Resick, P. A. (2014). Data mining in psychological treatment research: A primer on classification and regression trees. *Journal of Consulting and Clinical Psychology*, 82(5), 895.
- Lambert, M. J., Gregersen, A. T., & Burlingame, G. M. (2004). The outcome questionnaire-45. *The use of psychological testing for treatment planning and outcomes assessment*, 3, 191-234.

- Lambert, M. J. (2015). Progress feedback and the OQ-system: The past and the future. *Psychotherapy*, 52(4), 381.
- Leamy, M., Bird, V., Le Boutillier, C., Williams, J., & Slade, M. (2011). Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *The British Journal of Psychiatry*, 199(6), 445-452.
- Leserman, J., Whetten, K., Lowe, K., Stangl, D., Swartz, M. S., & Thielman, N. M. (2005). How trauma, recent stressful events, and PTSD affect functional health status and health utilization in HIV-infected patients in the south. *Psychosomatic medicine*, 67(3), 500-507.
- Muthén, L. K., & Muthén, B. O. (2014). Mplus (Version 7.3)[computer software].(1998-2014). *Los Angeles, CA: Muthén & Muthén.*
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*, 14(4), 535-569.
- Schnurr, P. P., Hayes, A. F., Lunney, C. A., McFall, M., & Uddo, M. (2006). Longitudinal analysis of the relationship between symptoms and quality of life in veterans treated for posttraumatic stress disorder. *Journal of consulting and clinical psychology*, 74(4), 707.
- Shin, D. C., & Johnson, D. M. (1978). Avowed happiness as an overall assessment of the quality of life. *Social indicators research*, 5(1-4), 475-492.
- Slade, K., Lambert, M. J., Harmon, S. C., Smart, D. W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. *Clinical psychology & psychotherapy*, 15(5), 287-303.
- Gudicha, D. W., Tekle, F. B., & Vermunt, J. K. (2016). Power and sample size computation for Wald tests in latent class models. *Journal of Classification*, 33(1), 30-51.
- Vander Zee, K. I., Sanderman, R., Heyink, J. W., & de Haes, H. (1996). Psychometric qualities of the RAND 36-Item Health Survey 1.0: a multidimensional measure of general health status. *International journal of behavioral medicine*, 3(2), 104.
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. *Applied latent class analysis*, 11, 89-106.
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons.
- Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). The ptsd checklist for dsm-5 (pcl-5). *Scale available from the National Center for PTSD at www.ptsd.va.gov.*
- Wells, M. G., Burlingame, G. M., Lambert, M. J., Hoag, M. J., & Hope, C. A. (1996). Conceptualization and measurement of patient change during psychotherapy: Development of the Outcome Questionnaire and Youth Outcome Questionnaire. *Psychotherapy: Theory, Research, Practice, Training*, 33(2), 275.
- Zatzick, D. F., Marmar, C. R., Weiss, D. S., Browner, W. S., Metzler, T. J., Golding, J. M., ... & Wells, K. B. (1997). Posttraumatic stress disorder and functioning and quality of life outcomes in a

nationally representative sample of male Vietnam veterans. *American Journal of Psychiatry*, 154(12), 1690-1695.