Title

Psychiatric Phenotyping Using Symptom Profiles: Can Self-Report Symptoms Inform a New Psychiatric Taxonomy?

Author

Ross, Jessica

Publication Date

2018

Peer reviewed|Thesis/dissertation

Psychiatric Phenotyping Using Symptom Profiles: Can Self-Report Symptoms Inform a New Psychiatric Taxonomy?

by

Jessica Ross

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree

DOCTOR OF PHILOSOPHY

in

Biological and Medical Informatics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

DEDICATION AND ACKNOWLEDGMENTS

This work is dedicated to my wonderful family, including my parents, Stuart and Stephanie Ross, Paul McIntire and David Ross. Without their continuous love and support, the completion of this undertaking would not have been possible.

I would like to thank the members of my committees. They all took time out of their extremely full schedules to advise and support me. First ,Dr. Ida Sim, for inspiring me and always pushing me forward, both intellectually and practically. Dr. Thomas Neylan has served as a mentor since residency, and has guided me through fellowship, and continued to do so through this work. Kevin Delucchi and Donna Hudson have always been available with advice and insight, and without them this work would not have been completed. Finally, Amar Das served on my proposal committee despite being in another institution, and has deeply affected my work as one of the leaders in the field of bioinformatics in the psychiatric domain.

Further, I want to thank several wonderful friends. First, Portia Kersten has been with me from before the beginning of this project, and I know she will be with me until after the end. Next, Cindy McSherry, Marisa Felt, and all of the CDS mothers who are the village that kept our family afloat throughout these wonderful and challenging years. Finally, all of my colleagues and friends at Zuckerberg San Francisco General, including Jonathan Nye, Alissa Peterson, Andrea Rosati, Emily Lee, Mark Leary, Jim Dilley, and all of the nurses, social workers, occupational therapists, and administrative staff on 7B an 7C. Thank you for challenging me and inspiring me every day with your devotion to our patient population. I sincerely hope my work will help lead to the improvement of care for those whom we serve.

Also, I would like to acknowledge The Veteran’s Administration Advanced Fellowship in Psychiatry/Neurosciences, who supported this research.

Psychiatric Phenotyping Using Symptom Profiles: Can Self-Report Symptoms Inform a New Psychiatric Taxonomy?

by

Jessica Ross

The Diagnostic and Statistical Manual (DSM) has served as the gold standard for psychiatric diagnosis for the past several decades in the United States, and it mirrors mental health and substance abuse diagnoses in the ICD-9 and ICD-10, which are used in numerous other countries. However, DSM diagnoses have severe limitations when used as phenotypes for studies of the pathophysiology underlying mental disorders, as well as for clinical treatment and research. This dissertation proposes a novel approach for deconstructing DSM diagnostic criteria using expert knowledge to inform feature selection for unsupervised machine learning. A multimodal dataset comprised of combat veterans, approximately one-third of whom had received a DSM-IV diagnosis of post-traumatic stress disorder (PTSD), is used in these analyses. Unsupervised learning methods are employed to identify robust groups of patients who clustered together with respect to clinical symptoms. Symptom profiles are used to stratify subjects into cohorts who have clinical and biological homogeneity, irrespective of their DSM diagnoses. Clusters identified suggest that prior contrasting biomarker findings in patients with PTSD may be due to heterogeneity that is reduced when using phenotypes derived from self-report psychiatric symptoms. Results of these analyses can be represented in rich clinical phenotypes

that relay both clinical and biological markers of interest. These findings suggest that itemized self-report symptom data may be useful to inform a new taxonomy for psychiatry, enhancing the bidirectional translation of knowledge from the bench to the clinic through a common terminology.

Key words: clinical and translational informatics, psychiatry, taxonomy, unsupervised learning, k-means, hierarchical cluster analysis

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

The *Diagnostic and Statistical Manual of Mental Illness, Fifth Edition* (American Psychiatric Association. & American Psychiatric Association. DSM-5 Task Force.) is the clinical psychiatric classification system for mental illness and substance abuse disorders currently used in the United States. When diagnosing using the DSM, the clinician reviews numerous symptoms with the patient to determine the symptoms' presence or absence, and then uses multiple DSM algorithms to determine the best diagnosis for the patient. DSM syndrome phenotypes are known to have severe limitations when used for identifying biomarkers associated with mental illness (Craddock & Owen, 2010; Cross-Disorder Group of the Psychiatric Genomics et al., 2013; Franklin & Zimmerman, 2001; Insel, 2014; Jablensky, 1999). It is unclear whether there is a more effective way to use the self-report symptom data obtained during the diagnostic process as phenotypes for determining the pathophysiology of mental illness. The purpose of this study is to determine if primary symptom data obtained by mental health practitioners can be used to develop phenotypes that effectively identify clinical psychiatric patient cohorts with increased pathophysiological homogeneity as compared to patients aggregated by DSM diagnosis. These data are obtained as part of the normal clinical workflow and are documented in patients' medical records. As these data are routinely ascertained through standard-of-care clinical psychiatric practice, they may theoretically be used to phenotype clinical populations, facilitate biomarker discovery based on cohort identification, and allow observational studies on patient outcomes in different cohorts. Ideally, this will lead to the identification of biomarkers that can be used to stratify psychiatric populations in the future, which will improve diagnosis, treatment, and outcomes in these populations.

Brief Review of the History of Psychiatric Classification

The DSM was developed after World War II, largely to monitor the prevalence of mental illness in soldiers who returned from the war and provide them with appropriate treatments. The DSM-I (1952) and DSM-II (1968) were both rooted in psychodynamic psychiatry, consistent with the legacy of psychoanalysis (e.g., beginning with Freudian theory) (American Psychiatric Association. Committee on Nomenclature and Statistics., 1952, 1968). The DSM-III (1980) took an "atheoretical" approach. That DSM used a classification system based on clustered patterns of symptoms determined by expert consensus in committee meetings, without the explicit consideration of empirical data or explanatory models of disorder etiologies (American Psychiatric Association. & American Psychiatric Association. Work Group to Revise DSM-III., 1987). This is called the Kraepelinian approach, in reference to Emil Kraepelin, the late 19th century psychiatrist known for classifying psychiatric patients through careful observation (Craddock & Owen, 2010). Studies performed after publication of the DSM-III verified that this new approach vastly improved interrater reliability with respect to patients' psychiatric diagnoses (Segal, Hersen, & Van Hasselt, 1994). This approach presumed that these categories reflected underlying pathological processes; however, the DSM-III did not include any quantitative data or biological markers, as none were identified that had strong enough associations to the DSM syndromes to have clinical utility (Hyman, 2010).

The DSM-IV was the gold standard in psychiatric treatment and research from 1994 to 2013 (American Psychiatric Association. & American Psychiatric Association. Task Force on DSM-IV., 1994). It was supplanted by the DSM-5, which was officially released in May 2013 (American Psychiatric Association. & American Psychiatric Association. DSM-5 Task Force.).

Limitations of the DSM

The limitations of the DSM regarding the use of its syndromes as phenotypes in clinical practice and research have been well documented (Craddock & Owen, 2010; Cuthbert, 2014; Insel, 2014; Maj, 2005, 2014). This review will focus on the most salient points as they relate to phenotypic heterogeneity and limitations of syndromes for identifying biomarkers, specifically on the details and structure of DSM-IV syndromes, as the analyses in this dissertation utilize data ascertained with DSM-IV phenotypes. However, the DSM-5 retains the same basic structure, and thereby limitations, of the DSM-IV (Insel, 2014). This discussion will delineate informatics challenges to the construction of phenotypes that define more clinically and biologically homogenous groups from symptom data obtained using the DSM-IV. These issues will highlight the importance of the studies conducted in this dissertation.

DSM-IV defines 137 syndromes/disorders across 15 categories, and as with DSM-5, can be conceptualized as a dichotomous polythetic classification system. The polythetic aspect means that a person can be diagnosed with a disorder with only a subset of the criteria that define the syndrome, while the dichotomous aspect delineates that an individual is either given the diagnosis or is not (Andreasen, 1995; Silverman, 2014). These qualities result in large phenotypic heterogeneity within DSM syndromes (Sharp et al., 2016). For example, with the diagnosis of PTSD in DSM-IV, there are 1,750 different groupings of symptoms that can lead to the diagnosis (Rosen, Lilienfeld, Frueh, McHugh, & Spitzer, 2010).

Furthermore, DSM-IV phenotypes have limited clinical utility, as two patients with the same DSM-IV syndrome may not share any of the same symptoms. For example, in the case of major depressive disorder (MDD), “depressed mood” or “anhedonia” and five out of nine other symptoms are required for diagnosis, leading to 112 different possible symptom presentations.

Because some of the symptoms are underspecified (e.g., sleep disturbance can refer to hypersomnia or insomnia), two patients can have MDD without sharing any of the same symptoms. This lack of syndromic specificity was carried over into the DSM-V, as many major syndromes, including MDD, were not updated in the new edition.

Symptom overlap across syndromes also results in clinical presentations in which patients are diagnosed with several comorbid psychiatric disorders (Hyman, 2010; Maj, 2005). For example, both MDD and PTSD present with sleep disturbances and feelings of guilt. As a result, diagnoses and treatments are complicated because of the paucity of evidence-based treatment algorithms for comorbid psychiatric disorders (Sharp et al., 2016).

Phenotypic heterogeneity also causes issues for clinical and translational research. Several classes of medications have been shown to be efficacious across several syndromes in clinical trials (Hales, Yudofsky, Gabbard, & American Psychiatric Publishing). Clinically, these medications are used to treat several DSM-IV classes of disorders (e.g., mood disorders and anxiety disorders, mood disorders and psychotic disorders). This overlap of symptoms and efficacy of pharmacological agents across multiple syndromes could indicate a common pathophysiology present across different DSM-IV syndromes (Hyman, 2010).

The lack of biological validity underlying the DSM-IV syndromes has hampered the identification of robust biomarkers within syndromes (Craddock & Owen, 2010). Consistent with the overlapping symptoms seen among DSM-IV syndromes, multiple studies have observed altered biological markers across syndromes. One recent study showed that calcium voltage-gated channel subunit alpha1 C (CACNA1C) variants were associated with several psychiatric syndromes, including autism spectrum disorder, attention deficit hyperactivity disorder (ADHD), bipolar affective disorder, and MDD (Cross-Disorder Group of the Psychiatric Genomics et al.,

2013). In addition, several other biological markers (e.g., altered suppression of cortisol by dexamethasone causing dysregulation of the hypothalamic–pituitary–adrenal [HPA] axis) have been shown to be significantly different in affected individuals than in controls in multiple psychiatric syndromes (Kirschbaum & Hellhammer, 1989).

Finally, the current terminology of the DSM nosology makes it difficult to investigate symptoms across syndromes because the DSM-III was based on the premise that no symptom could be replicated in two syndromes (Franklin & Zimmerman, 2001). It is unclear if this rule was explicitly carried over to the DSM-IV and DSM-V; however, as the precursor to these classification systems, we may presume its influence. This issue increases the challenges of using existing natural language processing (NLP) and text-mining methods to automatically identify symptoms across syndromes within narrative text or structured interview data. The APA has published several cross-cutting symptom measures that may be used to investigate primary psychiatric symptoms that are present across syndromes in clinical populations, although these measures are not routinely used clinically, and their use is not currently reflected in the literature (American Psychiatric Association. & American Psychiatric Association. DSM-5 Task Force.),

The Research Domain Criteria (RDoC)

Given the limitations of the DSM, major stakeholders in mental health treatment and research have publicly presented alternative systems for codifying current and future knowledge. The most notable of these systems is the psychiatric incarnation of Precision Medicine: the Research Domain Criteria (RDoC) (Insel, 2014). RDoC is focused on classifying mental disorders on dimensions of observable and neurobiological measures (Cuthbert, 2014; National

Research Council (U.S.). Committee on A Framework for Developing a New Taxonomy of Disease., 2011). While not explicitly excluding self-report symptoms, the RDoC Project is understandably focused on measurable phenomena, and may run the risk of being too distant from the clinical phenomena that will continue to lead patients to seek medical help for the foreseeable future (Cuthbert, 2014; Jablensky & Waters, 2014; Maj, 2014; Sharp et al., 2016).

It is not yet clear how RDoC will be implemented in a natural clinical setting (Cuthbert, 2014; Sharp et al., 2016). Nor is the process apparent of transforming from the current state of psychiatric clinical evaluation, where treatment relies solely on patient self-report and clinician-observed phenomena, to one that relies on biologically based constructs that lack social and clinical currency (Sharp et al., 2016(Cuthbert, 2014)). Presumably, as with other domains of medicine, a patient's presenting complaint will lead to an analysis of biological markers that will direct diagnoses and treatment. Therefore, the identification of symptoms and signs (as opposed to psychiatric syndromes) that may inform the differential diagnoses and subsequent workup may facilitate the translation of research findings into clinical practice and, hence, improved patient outcomes.

Furthermore, generations of psychiatrists have been trained using the DSM. They routinely use symptom-level data to evaluate patients and identify the most appropriate DSM disorder(s) for the patient based on the constellation of symptoms at clinical presentation. Additionally, remuneration of clinical services and many clinical and translational research grants still depend on the DSM classification system (Sharp et al., 2016).

Aims

For the reasons described, it is important to consider how and where patient self-report symptoms will be used in the psychiatric domain in the age of Precision Medicine. The primary aim of this project is to use mental health self-report symptom data to construct psychiatric symptom profiles that enable patients to be grouped together with increased clinical and biological homogeneity when compared with DSM syndromes as phenotypes. Furthermore, this study will identify biomarkers that differ significantly between these symptom profiles to demonstrate that self-report symptoms can be used to inform the development of a new taxonomy for psychiatry that is biologically valid while retaining clinical relevance. Finally, a graphical representation of the above findings is introduced as a rich clinical phenotype that includes both clinical and multimodal biological data. This phenotype facilitates clear visualization of markers that delineate clusters or subgroups of patients, and may potentially be used to begin to integrate the types of findings described in this study within the RDoC framework.

Chapter 2: Review of the Literature

Introduction

Numerous studies have attempted to clarify underlying relationships between individual psychiatric symptoms and psychiatric syndromes. Many studies used data reduction methods to identify the structure underlying numerous psychiatric symptoms measured in clinical evaluation and research studies. One commonly used feature reduction method is Principal Component Analysis(PCA), which identifies components that contribute to the variance observed in a set of variables, such as a group of psychiatric symptoms in a clinical inventory (Hastie, Tibshirani, & Friedman, 2009). Another frequently used method is exploratory factor analysis (EFA), which explicitly represents communalities during the extraction of a set of factors underlying a larger group of symptoms, to explain the relationships between these symptoms (Dazzi, Shafer, & Lauriola, 2016; Widaman, 2007).

In contrast to those studies, the goal of this study was to investigate if groups of individuals with increased clinical and biological homogeneity can be identified using psychiatric symptom data and an atheoretical data-driven approach rather than using the DSM syndrome as a phenotype. This study also sought to validate these symptom-level groupings using clinical and biological markers that were not used in previous cluster analyses. Given the current state of psychiatric diagnoses, where the number and qualities of these hypothesized homogenous groups are unknown, clustering algorithm analyses are an ideal method to investigate this question (Hastie, Tibshirani, & Friedman, 2001).

In this review of the literature, the strengths and weaknesses of model-based and non-model-based cluster analyses and the rationale for using non-model-based clustering algorithms for this study will be discussed. This review will focus on studies that implemented cluster

analyses using psychiatric symptom data as features to identify groupings as is proposed in this dissertation, and it will summarize the overall findings.

Latent Class Analysis

Several psychological and medical studies have been performed using both cluster analysis and latent class analysis (LCA), a form of finite mixture modeling (Goodyer, 2012). The primary difference between these methods is that LCA is model-based, whereas other types of cluster analysis (e.g., k-means and hierarchical) are not.

A model-based clustering approach allows one to incorporate the probability that an individual belongs to a class based on the statistical distribution of that class in the dataset. These methods assume conditional independence, i.e., all features in the analyses are assumed to lack direct relationships with each other and are only connected based on latent variables (Uebersax, 1999). The use of latent class methods with psychiatric symptoms has been criticized because it violates this assumption, although it is possible to allow for within-class correlations in the modeling (McCrea, 2013; B Muthen, 2001; B. Muthen, Asparouhov, & Rebollo, 2006; Muthén & Muthén, 2012; Uebersax, 1999). For instance, conditional independence cannot exist between two variables, such as “increased appetite” and “decreased appetite,” if both questions are asked within the same sample, as was done in one study (Sullivan, Kessler, & Kendler, 1998). To assume that there is conditional independence between symptoms of anxiety and depression, psychosis and depression, or insomnia and depression seems premature.

One of the most impressive studies to date in this field used all of the items assessed in the Structured Clinical Interview for DSM (SCID) as features in an LCA to show that mixture modeling was able to stratify individuals with schizophrenia into two groups that were associated

with two different alleles within the DTNB1 gene (Wessman et al., 2009). The two groups identified in that study were: a) individuals with psychosis, predominant mood symptoms, and intact cognition; and b) individuals with early-onset psychosis, higher-level positive and negative symptoms, and cognitive impairment. Interestingly, these groups have been well documented previously, and arguably these results could have been obtained using non-model-based cluster analyses as well (Bora, Yucel, & Pantelis, 2010). A recent in-depth analysis of LCA methods in patients with anxiety and depression raised issues regarding the inherent assumption of conditional independence. Furthermore, it showed that findings across studies were not consistent, indicating that LCA might not be ideal for these types of data (McCrea, 2013).

Cluster Analysis

In contrast to LCA, non-model-based clustering methods, such as hierarchical or k-means clustering (hereafter referred to as "cluster analysis"), have no inherent assumption of independence between features (Guyon & Elisseeff, 2003). Another benefit of cluster analysis is the ability to handle large amounts of variables with a relatively small subject sample (Hand & Heard, 2005). Heuristics have been developed to provide guidelines regarding the number of cases per feature that "should" be used in analyses; yet, there is no consensus on these guidelines. Moreover, many successful studies have disregarded these guidelines (Guyon & Elisseeff, 2003). Cluster analyses are often preceded by feature reduction, but the individual variables can also be left untransformed, e.g., in their symptom-level state (Hastie et al., 2009). This characteristic allows researchers to use cluster analysis to define symptom profiles with symptom data obtained through normal clinical practice, which may then be tested for different

associations with biomarkers or outcomes, and ultimately used for population stratification informing clinical practice.

The largest limitation in non-model-based clustering is arguably that these algorithms require the investigator to determine how many partitions to make in the data (Hastie et al., 2001). When the ideal number of partitions is not known *a priori,* the groupings identified by the analysis may not represent a "true" underlying structure. There have been multiple methods developed to test cluster validity, including a variety of indices that are used to determine cluster quality and are recognized as internal validation measures for cluster validity (Arbelaitz, Gurrutxaga, Muguerza, Perez, & Perona, 2013). Additionally, multiple studies also employ methods of external validation by showing differences between clusters using external measures of biological or clinical interest not included as features in the analyses (Everitt, 2011; Hastie et al., 2009). Both validation methods are used in the papers reviewed below as well as in the analyses described in this study.

The amounts and varieties of data gathered and stored in electronic health records is increasing, and it is relatively easy to gather self-reported symptom data through the internet. Thus, developing useful profiles of self-reported symptoms that can be used to stratify clinical populations for diagnoses and treatment is a desirable and potentially useful goal. Furthermore, profiles of self-report symptoms may help improve the diagnosis and treatment of mental illness by increasing homogeneity compared with current DSM diagnoses. Quantitative symptom profiles as psychiatric phenotypes are much more in line with the vision of Precision Medicine than binary DSM diagnoses. In addition, identifying and implementing methods to delineate psychiatric phenotypes that are less resource-intensive could facilitate large-scale identification of patients for cohort, longitudinal, observational, and interventional studies. Ultimately,

delineation of psychiatric phenotypes by quantitative symptom profiles could inform treatment algorithms and improve outcomes. Furthermore, the resultant aggregated data could be used for epidemiological studies to inform public health decisions relating to mental health issues (Wang et al., 2005).

Methods

A PubMed search for "psychiatric symptoms" and "Cluster Analysis" on September 23, 2016, produced 41 articles. Articles using psychiatric symptoms and psychiatric symptom subscales, alone or in conjunction with other clinical and demographic factors (but not biomarkers) in the cluster analysis, were included in this review. Several articles were excluded because they reported the clustering of neuropsychiatric or imaging measures instead of clinical psychiatric symptoms; one article was excluded because it was not available in English.

In addition, while many studies had clusters based on symptoms or combinations of symptoms and other data types (e.g., clinical, demographic), this review focused on studies conducting external validation of cluster differences. By design, a clustering algorithm will identify clusters within a dataset; nonetheless, the validity and utility of these partitions requires demonstrating differences in external factors not used in the clustering analyses (Everitt, 2011; Hastie et al., 2009). Only studies evaluating the validity of clusters using external factors were included, as the goal in the present study is to identify cohorts that differ with respect to biological and clinical outcomes not used as features in the cluster analyses. Ultimately, 14 articles were included in this review.

Results

Studies that met our inclusion criteria were grouped into three major categories. Studies in the first category compared the clusters identified to existing classification systems, predominantly the DSM. One study also compared findings to the diagnostic criteria for Gulf War Syndrome. The second category compared clusters across other clinical and demographic variables not included as features in the cluster analyses. The final category used psychiatric symptom-level data as features for cluster analyses to identify groups that were then tested for external validation using biological markers.

Beginning in the late 1960s, several studies described a variety of clustering algorithms for grouping individuals based on psychiatric symptoms to identify subtypes of major psychiatric disorders (Everitt, 2011). Although the motivation for these studies is not clear, the idea of "clustering" patients based on psychiatric symptoms appealed to behavioral health experts and the machine-learning community. During the same decade, the first atheoretical DSM (DSM-III) was published, which classified patients presenting with similar clusters of symptoms. Many studies attempted to identify subtypes of individuals based on available psychiatric symptom data. These studies classified depression into neurotic, endogenous, and psychotic subtypes (Paykel, 1971; Pilowsky, Levine, & Boulton, 1969); explained the heterogeneity in schizophrenia (Farmer, McGuffin, & Spitznagel, 1983); and described subgroups of eating disorders (Hay, Fairburn, & Doll, 1996). One of these early studies used hierarchical clustering of 39 psychiatric symptoms to identify four groups of patients with different clinical presentations who received different treatments with different efficacies. No equivalent differences were identified when the individuals were grouped by DSM diagnosis (Williams, Barton, White, & Won, 1976).

A more recent study using Hierarchical Cluster Analysis (HCA) to investigate symptoms of atypical depression across patients with unipolar and bipolar disorders found no difference in the depressive-symptoms profile between the two diagnostic groups. Five clusters produced the most stable solution. The only significant difference in external variables between the two clusters was the number of previous episodes of MDD. The cluster that had the most episodes of MDD also had higher means for all five symptoms used in the cluster analysis, implying greater overall depression severity in this group (Robertson et al., 1996). A different study in patients with schizophrenia used 55 psychiatric symptoms to determine if clusters identified correlated with the clinical subtypes of schizophrenia defined in the DSM. No concordance was found (Helmes & Landmark, 2003).

A large-scale investigation of more than 460,000 individuals who sought first-time mental health counseling in the New York City metropolitan area within 27 months after the September 11, 2001, attacks attempted to determine if these individuals reported symptom patterns that concurred with DSM syndromes. HCA of 31 self-reported symptoms was used to identify seven clusters of individuals. One cluster had symptoms strongly overlapping with those used to diagnose PTSD; another overlapped with MDD; and a third overlapped with comorbid MDD and PTSD. On the other hand, across 27 months of clinical data, more than 50 percent of individuals who sought counseling after the event did not fit into one of those three diagnostic groups or have symptoms consistent with any DSM diagnosis (Jackson et al., 2006).

A related effort evaluated whether a data-driven approach could identify clustered patterns of symptoms that would separate Gulf War veterans (GW) from non–Gulf War veterans. K-means clustering of sociodemographic factors, health variables, and scores from 10 symptom groups used in a sample of 500 veterans randomly selected from three U.K. military cohorts

identified five clusters as the optimal solution. The clusters differed in degree across nine out of 10 symptom groups. Three clusters overlapped in the intensity of the 10th group, musculoskeletal symptoms. The authors concluded that the results of this study did not support the existence of Gulf War Syndrome because there were no specific symptom clusters that could stratify GW veterans from veterans in other conflicts or those who were not deployed. Furthermore, with the exception of musculoskeletal symptoms, the five groups were stratified by the intensity of nine types of symptoms, with the most affected group having the highest level of all nine types of symptoms, and the least affected group having the lowest intensities across all nine types of symptoms(Everitt, Ismail, David, & Wessely, 2002).

The four studies discussed above did not identify further differences in external clinical, demographic, or biological markers across identified clusters. In contrast, the following studies investigated how diagnoses differed across identified clusters as well as demographic and clinical variables not initially used in the cluster analysis. These studies enrolled several different clinical populations, ranging from a primary care population to individuals hospitalized due to dementia. Psychiatric symptoms used in the cluster analyses included features such as personality measures, somatic symptoms, items on the Alcohol Withdrawal Scale, and the rating scales for mania and depression. The findings were generally divided into two groups: one based on symptom profiles that overlapped and therefore differed qualitatively, and the other based on symptom severity.

HCA was conducted on 11 summary scores from the Millon Clinical Multiaxial Inventory (MCMI) in 137 subjects diagnosed with obsessive-compulsive disorder (OCD). Four clusters were identified: those with no evidence of personality pathology, those with dependent and compulsive pathology who had the best treatment adherence and outcomes, those

who could be classified as histrionic/borderline, and those with schizoid, dependent, schizotypal, and avoidant interpersonal issues. This study concluded that different personality profiles had relevance in treatment outcomes of patients with OCD (Fals-Stewart & Lucente, 1993).

In 96 males hospitalized for bipolar disorder or manic episodes, 19 features selected from the Young Mania Rating Scale (YMRS), Montgomery-Asberg Depression Rating Scale (MADRS), and Scale for Assessment of Positive Symptoms (SAPS) were used to conduct a factor analysis followed by HCA of the three factors. This analysis resulted in two distinct clusters: Cluster 1 had higher psychomotor elevation and Cluster 2 had higher psychotic symptoms and depression. Cluster 2 had higher overall substance-use disorders, which was identified using external data not included in the initial analysis (Guclu, Senormanci, Aydin, Erkiran, & Kokturk, 2015).

Neuro-cognitive and socio-cognitive measures from 100 individuals with anorexia nervosa were subjected to HCA. The analyses defined three clusters with different levels of overall functioning. There were no differences across the clusters in factors not used in the analysis, however, including clinical characteristics, service utilization, or treatment adherence (Renwick et al., 2015).

One study investigated nine items of a diagnostic tool developed to assess mental health and social functioning of hospitalized patients with dementia using k-means and HCA. Four identified clusters differed in external measures of cognitive status, length of hospital stay, and legal admission status. The four clusters, termed affective, functional, somatic, and psychotic, were based on “clinical meaningfulness” with affective and functional denoting less severe cognitive and functional impairment than somatic and psychotic clusters. The significance of differences between clusters was not clear as no post-hoc statistical tests were performed on the

variables used for external validation. In addition, because 25 to 46 percent of the data were missing for the Mini Mental Status Exam across the four clusters, severity could not be interpreted across clusters (Ortoleva Bucher, Dubuc, von Gunten, Trottier, & Morin, 2016).

To delineate clinically meaningful clusters of somatic symptoms in a primary care population, 1,466 primary care individuals were evaluated using a panel of 40 self-reported symptoms to identify unexplained medical complaints. Initially, the authors used grade of membership method to reduce the features from 52 to seven symptom groups. The analysis identified one cluster with significantly greater levels of psychiatric morbidity, functional impairment, the widest variety of somatic symptoms, and a different demographic profile than the other clusters. Despite these findings, there were no significant differences observed in external factors not used in the HCA among the other clusters. The authors concluded that the finding supported delineation of multisystem somatic symptom comorbidity as a proxy for overall somatization severity in this population (Gara, Silver, Escobar, Holman, & Waitzkin, 1998).

A study of 207 individuals going through alcohol withdrawal measured 17 items on the Alcohol Withdrawal Scale (AWS) and analyzed the data using HCA. Ten of the items were considered vegetative symptoms, and seven were considered psychopathological symptoms. The authors determined that a five-cluster solution was optimal. The five clusters differed in the severity of AWS score, with the most severe clusters having 100 percent delirium tremens with hallucinations, and significantly higher psychopathological symptoms than the lower cluster. Regarding variables not used in the analyses, the only significant difference between clusters was in age; however, post-hoc differences between clusters were not reported and therefore

significant differences between clusters were not interpretable (Driessen, Lange, Junghanns, & Wetterling, 2005).

Finally, an HCA study of 1,788 healthy Chinese college students used responses to the three subscales of the 20-item Toronto Alexithymia scale as features, and identified four clusters, three of which were identified as individuals having alexithymia. One cluster, the high-alexithymic group, had the highest scores across all subscales, whereas individuals in the two other alexithymic clusters were classified as alexithymic introverts with increased difficulty identifying feelings and alexithymic extroverts with an externally oriented cognitive style. External validation in this study was performed across several self-reported clinical measures that were not included in the cluster analysis. Validation showed that the highest alexithymic cluster scored significantly worse than the non-alexithymic group on all measures. There were no significant differences between the two less severe alexithymic clusters (Chen, Xu, Jing, & Chan, 2011).

Only three studies identified in this review investigated external biomarker differences across clusters defined using psychiatric symptoms alone as features. One study used 29 tic symptoms to perform HCA on 89 probands in families with Gilles de la Tourette syndrome. Eleven symptom clusters were identified that were then used to inform a Principal Component Analysis (PCA). Four factors were delineated, and they accounted for 60 percent of the variance in symptom presentation, comorbidity, recurrence risks, and within-family correlation. These factors may represent heritable components of Tourette's. The more severe symptoms (aggressive and compulsive factors) were associated with earlier age of onset and ADHD diagnosis in probands, whereas the less severe symptoms had no relation to external

characteristics not used in the analyses except for the association of simple tics with males (Alsobrook & Pauls, 2002).

Another study successfully delineated two distinct symptom clusters through HCA of 38 lifetime tic and related symptoms in two isolated populations with Tourette's. The clusters differed across ancillary data not used in the cluster analyses, including age of onset, medication treatment, and family history. Cluster 1 was identified as those with "simple tics," and Cluster 2 with "severe tics," as well as being associated with significantly more psychiatric comorbidity and poorer outcomes (Mathews et al., 2007).

Finally, a study of 332 males between the ages of 18 and 60 with a diagnosis of alcohol dependence used features from the Severity of Alcohol Dependence (SADD) scale and the Hamilton Depression Scale (HDRS) as well as other clinical factors, such as family alcoholism and problem-drinking onset age, to perform k-means clustering (Baltieri & Correa Filho, 2012). The authors delineated two clusters: Cluster 1 with a higher mean HDRS and Cluster 2 with a history of severe individual and familial alcoholism and a lower mean HDRS. Validation with data not used in the cluster analysis showed that Cluster 2 had higher plasma ALT and a lower chance of continuing treatment than Cluster 1. These results can be interpreted as two groups of individuals with alcohol dependence, one with a family history of alcohol dependence, as well as a more severe alcohol dependence, and the other with alcohol use related to depressed mood; however, these results were difficult to interpret as the exact features used for clustering were not documented.

The reviews above demonstrated two major categories of findings using cluster analyses. The first showed clusters that differed qualitatively, e.g., higher in some types of symptoms and lower in others. The second most prevalent finding was that clusters differed in overall severity

across all or most symptoms measured. This finding is interesting as it is not consistent with the implicit assumptions in the current DSM, that patients will present with patterns of symptoms that differ primarily qualitatively. However, it does lend support to claims that individuals with comorbid disorders may not really have multiple disorders, but instead have one underlying pathology that encompasses a multitude of symptoms in patterns not defined by the current psychiatric nosology (Hyman, 2010).

Chapter 3: Datasets Used in This Study

We conducted secondary analyses on patient data obtained through two studies conducted by Veterans Administration (VA) and Department of Defense (DoD) grants awarded to the San Francisco Veterans Administration Health Center. Details for both of these studies have been previously published (Apfel et al.; Samuelson, 2011; Samuelson et al., 2006; Schuff et al., 2008). The study design and sampling process are described briefly below.

Samples

The first study was conducted among Gulf War (GW) veterans to evaluate the neurological sequelae of Gulf War Illness (GWI). GW subjects (n=292) were recruited through the San Francisco Veterans advertisement and from a registry of GW veterans in Northern California, which was supplied by the study sponsor, the DoD. Study participants provided consent in accordance with the procedures approved by the Committee of Human Research at the University of California, San Francisco (Apfel et al., 2011).

The second dataset was ascertained through the Sierra Pacific Mental Illness Research, Education, and Clinical Centers (MIRECC), which collected data from 128 veterans, and can be separated into four groups according to two characteristics: with or without PTSD (PTSD+ or PTSD-, respectively), and with or without a history of alcohol abuse (ETOH+ and ETOH-, respectively). Thirty participants were PTSD+/ETOH+; 37 were PTSD+/ETOH-; 30 were PTSD-/ETOH+; and 31 were PTSD-/ETOH-. This dataset was used to determine whether volumetric and metabolic abnormalities in the hippocampus and other brain regions were present in PTSD, independent of alcohol abuse (Samuelson et al., 2006; Schuff et al., 2008). A

composite summary of demographic and clinical details of all male participants in both the GW and MIRECC datasets are displayed in Table 3.1. Only males, who accounted for 78 percent of subjects in these data, were used in these analyses to reduce heterogeneity of subjects given the relatively small sample sizes with respect to the analytic methods.

Table 3.1. Demographics of the Gulf War and MIRECC Datasets

Clinical Variables	**Gulf War**	**MIRECC**	**Combined**
All Participants	292	130	422
Males with Full Clinical Data	238	92	330
Age_{Male} (yrs) Mean (STD)	44.4 (8.8)	48.2 (9.3)	45.4 (9.3)
$Education_{Male}$ (yrs) Mean (STD	14.6 (2.3)	14.4 (2.3)	14.5 (2.2)
Trauma-Exposed $_{Male}$ (Meets Criterion A)	153	91	244
PTSD Diagnosis (DSM-IV)$_{Male}$	33	57	90
CAPS $Current_{Male}$ Mean (STD)	16.5 (24)	42.1 (34)	23.6 (29)
Current Alcohol Abuse (DSM-IV)$_{Male}$	88	38	126
Current Alcohol Dependence (DSM-IV)$_{Male}$	60	46	106
Current Major Depressive Disorder (DSM-IV)$_{Male}$	26	21	47
Lifetime Major Depressive Disorder (DSM-IV)$_{Male}$	100	65	165
Childhood Trauma (LSC) $_{Male}$	57	23	80

CAPS: Clinician-Administered PTSD Scale; LSC: Lifetime Stressor Checklist; MIRECC: Mental Illness Research, Education, and Clinical Centers study data

Inclusion and Exclusion Criteria

U.S. veterans or active-duty service members who had been deployed in wartime service were eligible to participate in this study. Subjects who met one or more of the following criteria were excluded from all studies: 1) history of head trauma, prolonged loss of consciousness (>10 min), neurological disorder, or systemic illness affecting central nervous system function (including all neurological disorders and diabetes); 2) current or previous history of any psychiatric disorder with psychotic features, presence of prominent suicidal or homicidal ideation, or use of antipsychotic medications during the past six weeks; 3) current major depression diagnosed using DSM-IV; 4) claustrophobia severe enough to prevent magnetic resonance imaging (MRI)/MRSI studies; 5) substantial concern that the sound of the MRI would evoke wartime memories and panic; and 6) presence of ferrometallic objects in the body that would prevent MRI/MRSI studies.

Sample Data

Several clinical and biological measures were obtained during sample collection, as listed in Table 2 and described in detail below. Among the collected data were clinical symptom data, and four different modalities of biomarker data. Table 3.2 presents a summary of these data types. Full details of the clinical interviews and self-report data are in the appendices.

Clinical Measures

The first assessment used a Structured Clinical Interview for DSM-IV (SCID) to identify exclusionary DSM-IV disorders and the Clinician-Administered PTSD Scale (CAPS) to identify

Table 3.2. Summary of Clinical and Biological Data Types Used

	Number of Items	Data Type	GW	MIRECC
Clinical Data				
Clinician Assessment of PTSD Symptom Assessment Questions	48	Likert	Yes	Yes
Clinician Assessment of PTSD Age of Trauma Questions	2	Continuous	Yes	Yes
Clinician Assessment of PTSD Summary Score	1	Binary	Yes	Yes
Beck Depression Inventory Symptoms Assessment Questions	22	Likert	Yes	No
Hamilton Depression Inventory Summary Score	1	Binary	Yes	Yes
Symptom Checklist-90 Symptom Assessment Questions	90	Likert	Yes	No
Symptom Checklist-90 Symptom Summary Scores	10	Continuous	Yes	Yes
Lifetime Drinking History Interview	3	Continuous	Yes	Yes
Structured Clinical Inventory for DSM-IV Alcohol and Substance Use, Abuse, and Dependence Scores	22	Likert	Yes	Yes
Lifetime Stressor Checklist	1	Binary	Yes	Yes
Neuropsychiatric Data				
CVLT	12	Continuous	Yes	Yes
WAIS	10	Continuous	Yes	Yes
Imaging Data				
Bilateral Frontal Lobe GM, WM, Sulcal CSF	6	Continuous	Yes	Yes
Bilateral Occipital Lobe GM, WM, Sulcal CSF	6	Continuous	Yes	Yes
Bilateral Temporal Lobe GM, WM, Sulcal CSF	6	Continuous	Yes	Yes
Bilateral Parietal Lobe GM, WM, Sulcal CSF	6	Continuous	Yes	Yes
Brainstem GM, WM, Sulcal CSF	6	Continuous	Yes	No
Cerebellar GM, WM, Sulcal CSF	6	Continuous	Yes	No
Subcortical CSF	1	Continuous	Yes	Yes
Bilateral Thalamus WM, GM, CSF	6	Continuous	Yes	No
Bilateral Caudate WM, GM, CSF	6	Continuous	Yes	No
Bilateral Lenticular Nuclei WM, GM, CSF	6	Continuous	Yes	No
Bilateral Hippocampal Volume	2	Continuous	Yes	Yes
Intracranial Volume	1	Continuous	Yes	Yes
Endocrine Data				
Cortisol in Area Under the Curve Pre (Day 1) and Post (Day 2) DST	2	Continuous	Yes	Yes
Genotype Data				
FKBP5/rs1360780 Alleles	2	Nominal	Yes	No
BDNF/Val66Met Alleles	2	Nominal	Yes	No
ApoE Alleles	2	Nominal	Yes	Yes

GW, Gulf War study data; MIRECC: Mental Illness Research, Education, and Clinical Centers study data; CVLT: California Verbal Learning Test; WAIS: Wechsler Adult Intelligence Scale; GM: gray matter; WM: white matter; CSF: cerebrospinal fluid; DST: Dexamethasone Suppression Test; FKBP5: FK506 binding protein 5; BDNF: brain-derived neurotrophic factor; ApoE: apolipoprotein E

post-traumatic stress disorder (Blake et al., 1995; First, 2002). The CAPS has 50 Likert scale items that are used to categorically diagnose individuals with PTSD. The Symptom Checklist-90-Revised (SCL-90-R) is a standard self-reported measure of general psychopathology, scored on nine primary dimensions and three summary indices (Derogatis, 1975). The Beck Depression Inventory (BDI) (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) is a widely used self-report test that includes an inventory of 21 Likert scale items that a subject answers. The BDI is used as a measure of depression for statistical analyses. The Hamilton Depression Rating Scale (HAM-D) is another extremely well-validated measure of depression that provides a single continuous numeric score based on 21 Likert scale items (Hamilton, 1960). The HAM-D score was obtained in both the GW and MIRECC studies, so was used when the studies were compared and pooled.

Neuropsychiatric Measures

The battery used in each of the above studies assessed three domains of cognitive functioning: verbal memory and learning; visual memory and visual-spatial skills; and attention, working memory, and intellectual functioning. Detailed descriptions of all tests used in the battery have been described elsewhere (Shiino et al., 1993) and are reviewed briefly below. First, to assess verbal memory and learning, we used three variables from the California Verbal Learning Test (Delis, Freeland, Kramer, & Kaplan, 1988): the Trial 1 score, the Total Trials 1–5 score, and the Long-Delay Free Recall. Participants also completed the Logical Memory I and II subtests of the Wechsler Memory Scale, third edition (WMS-III). Second, to assess short- and long-term visual memory and visual-spatial skills, we used the Visual Immediate Index and Visual Delayed Index scores of the WMS-III. Third, working memory and attention were assessed using the Letter Number Sequencing, Spatial Span, and Digit Span subtests of the

WMS-III. Participants were administered several subtests of the Wechsler Adult Intelligence Scale, third edition (WAIS-III), and the vocabulary subtest score was used as an estimate of intellectual functioning. Testing took approximately two hours, including a 15-minute midsession break. The participants were instructed to abstain from using alcoholic beverages and were breathalyzed before neuropsychological testing. Participants also provided urine for urinalysis for drug toxicology on the day of the neuropsychological assessment.

Neuroendocrine Measures

Because there have been several observations that subjects with PTSD have low resting cortisol levels and impaired HPA feedback responses (Yehuda, Boisoneau, Lowy, & Giller, 1995), a low-dose dexamethasone suppression test (DST) was performed. Subjects completed a low-dose (0.5 mg) dexamethasone (DEX) suppression challenge (Yehuda et al., 1995) to detect cortisol in saliva samples, which were collected and subsequently stored at -70°C (Kirschbaum & Hellhammer, 1989) and measured by the Clinical Laboratory at the San Francisco Veterans Affairs Medical Center (SFVAMC). Cortisol was measured on two consecutive days. First, serum cortisol was measured on Day 1, and at four consecutive time points 30 minutes apart after waking. The dexamethasone was then administered at 11 p.m. of Day 1, and on Day 2, the same four time points were measured as those on Day 1. The area under the curve (AUC) was calculated for each day's total, as well as for each 30-minute time period of each day. The resting cortisol level and the values for percent suppression between Day 1 and Day 2 were use in these analyses. Salivary dexamethasone levels were also measured during the first-time period on Day 2 to use as a covariate in ANCOVA analyses.

Imaging Measures

Structural MRI data were acquired using a 1.5 Tesla scanner (Vision, Siemens Medical Systems, Iselin, NJ) and a 3-D magnetization prepared T1-weighted gradient echo sequence (MPRAGE) with the following parameters: repetition time/spin-echo time/inversion time =10/4/300ms, 1x1 mm^2 in-plane resolution and 1.5 mm slab thickness, angulated perpendicular to the long axis of the hippocampus. These methods have been described in detail in previous publications (Apfel et al., 2011; Schuff et al., 2008).

Semiautomated hippocampal volumetry was carried out as described in detail previously (Hsu et al., 2002), using a commercially available high dimensional brain mapping tool (Medtronic Surgical Navigation Technologies, Louisville, CO) that has been validated and compared to manual tracing of the hippocampus (Hsu et al., 2002). Briefly, measurement of hippocampal volume is achieved first by manually placing 22 control points as local landmarks for the hippocampus on individual brain MRI data, and second, by applying fluid image transformations to match the individual brains to a template brain (Christensen, Joshi, & Miller, 1997). The pixels corresponding to the hippocampus are then labeled and counted to obtain volumes. This method of hippocampal volume measurement has a documented reliability of an intra-class coefficient better than 0.94 (Hsu et al., 2002). Intracranial volume was determined with Freesurfer, which uses an atlas-based spatial normalization procedure on T1-weighted images (Buckner et al., 2004).

Genetic Data

Genomic DNA was extracted from whole blood using the Promega Wizard Genomic DNA Purification Kit (Promega Biosystems, Sunnyvale, CA). Samples were genotyped at the

University of California Genomics Core Facility using an ABI 3730xl DNA analyzer (Applied Biosystems Inc., Foster City, CA). Sequencer DNA Sequence Analysis Software (Gene Codes Corporation, Ann Arbor, MI) was used to analyze the FKBP5 rs1360780 alleles, the ApoE4 alleles, and the BDNF Val66Met alleles.

Data Preparation

Initially, data were reviewed and recoded for logical missing values. For example, in questionnaires where a "No" answer to a question entailed skipping the next group of questions, subsequent values were not considered missing. For each analysis, rows and columns in which >75 percent of the data were missing were removed. The remaining missing values in ordinal data were labeled as "Missing," and continuous and Likert data were median-imputed. Outliers were investigated using scatterplots to determine if the variables were entered erroneously, and those determined to be erroneous were replaced with median-imputed values of all of the continuous or Likert values in the dataset for the variable. Less than 1 percent of the symptom, neuropsychiatric, and endocrine data required median imputation. No imputation was performed on imaging data, as all individuals with >75 percent of imaging data present had full imaging datasets. All data used in the clustering analyses were standardized and normalized. Table 3.2 summarizes the data types.

Complexity of the Dataset

The dataset described is notable for its clinical complexity across subjects. While in general this level of clinical heterogeneity across subjects is viewed as a weakness in biological studies, in this investigation the range of psychiatric disorders was conceptualized as a strength.

The aim of this study was to determine if psychiatric symptoms that were present across individuals with different DSM diagnoses can be utilized to identify groups that have greater clinical and biological homogeneity than those stratified by DSM diagnoses. As such, using a dataset with a very homogeneous and clean group of cases without any comorbidities (in this case, PTSD) and controls would greatly reduce the ability to determine how symptoms that are present across syndromes correlate with clinical and biological markers. As the Gulf War dataset was ascertained as a convenience sample to look into phenomena not specifically related to PTSD, the dataset has a richness in the amount and granularity of psychiatric symptom ascertainment that is often difficult to obtain. The dataset is also rich in psychiatric history and biomarkers, including biomarkers that have reports of conflicting correlations with DSM diagnoses in the literature, such as hippocampal size and cortisol suppression in individuals with PTSD as compared with controls (Pitman et al., 2012). Thus, as the hypothesis underlying this study, that heterogeneity within individuals with PTSD may to some degree explain these inconsistent reports, the described overall heterogeneity of the Gulf War dataset was considered an asset.

The MIRECC dataset was ascertained to investigate the specific hypotheses as to whether hippocampal abnormalities found in individuals with PTSD are independent of alcohol use. Unfortunately, due to the precision of this hypothesis, much of the granularity of the symptoms within the clinical inventories was not retained in the dataset. The dataset, however, was ascertained using the same protocols at the same facility during the same time period as the Gulf War study. Therefore, combining the features that were available in both datasets for analyses was determined to be a worthy undertaking due to the increased power obtained with the larger

dataset, which retained a more modest degree of symptom heterogeneity, even though a large degree of symptom granularity was lost.

Chapter 4: Methodology and Workflow

Choice of Unsupervised Learning Methods

The hypothesis underlying this study is that there are groups of psychiatric patients who are homogeneous with respect to clinical symptoms and biological markers, and that these groups can be identified more accurately by stratifying patients using data-driven phenotypes derived from psychiatric symptoms as opposed to DSM diagnoses. Hierarchical and k-means clustering, which are unsupervised machine-learning methods that offer a purely data-driven approach to identify structure within a dataset, were the methods of choice to validate the premise. Additionally, unsupervised learning can be used when both the number of groups and the categories of groups within a dataset are not known *a priori*, as is the case in this study. As discussed in the literature review, earlier scholars had success using both model-based (LCA) and non-model-based clustering methods (k-means, hierarchical) with psychiatric symptom data. Model-based clustering assumes that no direct relationship exists between the observed variables used to derive the clusters, in this case, the psychiatric symptoms. Instead, the variables are related only by unmeasured (e.g., latent) constructs. An analytical design where symptoms can have a direct effect on each other (e.g., anxiety can have a direct effect on insomnia, and intrusive thoughts can have a direct effect on mood) is most representative of the experience of psychiatric patients. Thus, non-model-based clustering is utilized in the following analyses.

Choice of Initial Dataset

As shown in Tables 3.1 and 3.2, the GW dataset was a larger dataset with more granular clinical, imaging, and genetic data. Therefore, the initial goal was to maximize the use of these data and begin using the entire available GW dataset. Initially, the plan was to show that

subgroups in patient populations that carry a diagnosis of PTSD are both clinically and biologically heterogeneous, which may help to explain the relative lack of reproducibility of biological findings associated with PTSD in the literature. There were several reasons, however, that all individuals within the dataset were used in the analyses, not just the individuals with PTSD. First, the overall hypothesis was that the psychiatric symptoms observed regularly in clinical work and research studies occur not only in a continuous spectrum of individuals with mental health distress who seek treatment, but also in healthy controls (albeit to a lesser extent on average). Therefore, it made the most sense to look for patterns of symptoms across all individuals and not just those who were designated by DSM criteria as having a specific diagnosis. Second, psychiatric symptoms are present across syndromes, putatively to some extent because of common pathways that lead to the presentation of similar symptoms in individuals with different DSM diagnoses. Thus, it was imperative to include individuals who had other DSM diagnoses with overlapping symptoms (e.g., those with MDD) in these analyses, to identify biomarkers associated with symptom profiles instead of specific disorders. Furthermore, as discussed, the binary nature of the DSM does not enable the categorization of subclinical symptoms; thus, an individual with subclinical PTSD may be very similar to one who meets the criteria for the diagnosis of PTSD in presentation, perhaps with one less distressing symptom. There is no way to represent these differences with the DSM; thus, a patient who just barely met the criteria for PTSD would be grouped in a study with a patient who had severe PTSD, and the subclinical patient would be identified as a healthy control. Therefore, the results using all subjects could potentially provide insight into which clinical and biological markers can be used to classify patients within a dimensional taxonomy, as opposed to the existing binary classification system used. In addition, given the small effect sizes and lack of robust findings in

the literature when patients with PTSD were compared to unaffected individuals, we hypothesized that we would need to leverage our entire dataset to identify biomarkers that differed across cluster groups to approach or meet a level of significance. Within the GW study, there were only 33 individuals out of 238 men with full psychiatric symptom data who met the diagnosis for PTSD. Using only these 33 subjects and given the demographic statistical covariates necessary to compare biomarkers, these subgroups would have been too small given the expected effect sizes for comparisons needed to establish statistical significance. Therefore, all subjects in both datasets were used in these analyses.

Data Analyses

The overall workflow used in these analyses is shown in Figure 4.1, and described in detail below. There were four groups of data that were analyzed: the complete Gulf War dataset (GW), the Gulf War Summary dataset (GS) with only full CAPS and summary scores for the HAM-D and SCL-90 (the features also present in the MIRECC), the MIRECC Summary dataset (MS), and a Combined dataset with both the Gulf War and MIRECC Summary data (CS). Feature selection was used with the clinical inventories in the complete Gulf War dataset as described below. All symptom data was used for the other GS, MS, and CS datasets.

Feature Selection

The Structured Interview for DSM (SCID) has been used widely for psychiatric phenotyping in research. Ideally, we would have been able to use symptom data from the SCID, the tool used to identify DSM diagnoses in this dataset initially. However, only composite-level

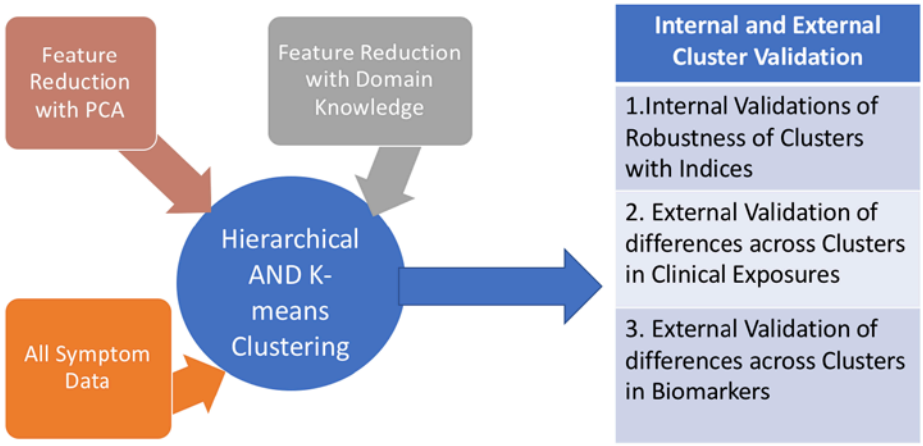

Figure 4.1. Overview of Workflow Used in these Analyses

SCID syndrome data was available in the datasets. In order to obtain an approximation as to the overlap between the questions on the SCID inventory and the complete clinical data available in the GW dataset, a domain expert (JR) mapped all of the items in the CAPS, BDI, and SCL-90 to SCID questions. The GW clinical data mapped to approximately 85 percent of all symptoms obtained in a formal SCID (data not shown). The major categories of missing data were psychotic symptoms (75 percent not represented in GW clinical inventories) and OCD symptoms (50 percent not represented in GW clinical inventories). As individuals with a history of psychosis were excluded from the GW and MIRECC studies, it was presumed that there was a broad representation of symptoms used for routine clinical psychiatric evaluation and diagnosis of nonpsychotic subjects ascertained in GW subjects.

The full set of psychiatric symptoms available in the GW dataset comprised data on those symptoms collected through the Clinician Assessment of PTSD (CAPS), the BDI, and the Symptom Checklist-90. These totaled 161 items, for which we had >99 percent full datasets from 238 male subjects (only three of these values were missing and therefore were median-imputed).

A common heuristic for identifying the optimal number of individuals to features in a cluster analysis is to include 10 individuals for each feature (Everitt, 2011); thus, the target number of items for our dataset was 29. Additionally, although cluster analyses have been successfully conducted with much less favorable ratios of cases to features (Iwao et al., 2002), removing features that are redundant is an often-favored way to potentially make findings from the analysis easier to understand and interpret by the end-users (Hastie et al., 2009). On the other hand, k-means clustering is not negatively affected by collinearity or dependence between variables, and there is no consensus on the optimal methodology for feature reduction (Carsten F. Dormann et al., 2012).

Based on this literature, three different approaches to feature selection were chosen, with the intent of performing the cluster analyses on the three different datasets and determining which cluster solutions were the most robust, and therefore likely to represent true underlying structure. In the first method, all of the items from the three clinical inventories were mapped by a domain expert (JR) to an expert list of symptoms developed by the APA to guide diagnoses with DSM-IV. We then compared items that mapped the same APA "higher" level symptom to a correlation matrix to affirm that both items represented the same clinical psychiatric phenomena (Appendices C1, C2, and C3). Clinical items were mapped to the APA symptom list if a clinical expert (JR) determined that the concepts identified were clinically equivalent (e.g., expected to result in the same answer if both questions were asked of the same patient in a clinical interview). A correlation matrix also was created for all clinical features to identify items with a correlation coefficient of ≥ 0.70. Fifty-four clinical items that mapped to the same APA symptom and had a correlation coefficient ≥ 0.70 were removed. While this approach did significantly reduce the clinical features, it should be noted that there was no consensus on how to determine

which highly correlated features should be removed. For example, if A and B have a correlation coefficient of 0.9, should feature A or feature B be removed? For this study, the initial item that the research subject answered was kept and the subsequent item was deleted (i.e., questions earlier in the inventory were kept, while later questions that correlated were deleted).

The second approach was Principal Components Analysis (PCA), a widely used statistical data-reduction method, to identify components that account for the majority of the variance found across all clinical items and transform them into a lower dimensional space (James, Witten, Hastie, & Tibshirani, 2013). The 161 GW clinical items required 40 principal components to explain 85 percent of the variance in the data, a threshold commonly used in similar studies (Hastie et al., 2001).

For the third approach, all of the available clinical variables from the CAPS, BDI, and SCL-90SCL-90 were used in the cluster analyses. This method was desirable because it did not reduce the data to constructs that were difficult to interpret clinically (as in PCA). It also has the benefit of identifying specific patient-reported symptoms that can effectively stratify clinical populations, thereby guiding the development of more focused clinical inventories.

Cluster Analysis

Two well-validated, unsupervised learning algorithms were used: k-means and HCA. Both algorithms have been widely used in a large variety of research domains for decades, including studies with datasets in the psychiatric symptom domain (Hastie et al., 2001). All statistical analyses were performed using R statistical programming software (Team, 2011). We performed both agglomerative hierarchical clustering (*hclust*) with Ward's distance measurement, which is a deterministic clustering method that will produce identical results every

time the analysis is conducted, and k-means nearest neighbor clustering (*kmeans*), which is nondeterministic and may produce different clustering solutions across several analyses (Hastie et al., 2001). Because these methods require the optimal number of clusters to be chosen by the analyst, and there is no absolute method to determine the optimal number of clusters, we used several well-validated indices, described below, to determine the optimal number of clusters, as well as the stability of the clusters created (Bayati, Davoudi, & Fatemizadeh, 2008; Lisboa, Etchells, Jarman, & Chambers, 2013).

The first method is the Calinski-Harabasz (CH) Index, which is also known as the variance ratio criterion. It is determined by the ratio of the overall between-cluster variance to the overall within-cluster variance (Caliński & Harabasz, 1974). The CH Index reaches its maximum when the between-cluster variance is relatively large and the within-cluster variance is relatively small. The second method for determining the optimal number of clusters was the lowest total sum of squares (wss) (Lisboa et al., 2013). For each cluster, the sum of squares is the sum of the distance between each point and the cluster centroid squared. Hence, the total sum of squares is the total variance of the observations, or the sum of the sum of squares for the entire clustering solution. This measure decreases as the number of clusters increases and the clusters become more homogenous. The optimal number of clusters is the point at which the rate of decrease in the total sum of squares decreases; in graphical form, this value appears as an “elbow” in the graph, and the number of clusters on the x-axis where the elbow occurs is considered optimal (Everitt, 2011). Both of these measures were plotted for the number of clusters k=2 through k=10, and the results were used to determine the optimal number of clusters for each solution. The results of these two indices never overlapped in this sample, which resulted in the analyses consisting of several cluster solutions for each dataset.

Both k-means clustering and HCA were implemented using the range of optimal clusters identified through the indices above. The Adjusted Rand Index (ARI) determines the similarity between two clustering solutions, where a higher ARI shows more similarity (Hubert, 1985). The ARI was selected to be one measure of robustness or underlying structure, as the structure was robust to different clustering methods. This workflow has been successfully used before to validate cluster robustness with psychiatric symptom data (Reser, Allott, Killackey, Farhall, & Cotton, 2015). Additionally, to increase the likelihood of finding true underlying data structure through clustering, bootstrapping with replacement was also employed to determine the robustness of the k-means clustering solutions and tracked the number of times each cluster dissolved out of 1,000 replications(Efron & Tibshirani, 1993) . This analysis was important as the sample size was relatively small, and the k-means clustering algorithm is nondeterministic. Results of the ARI and bootstrap analyses were used to identify the most robust clusters for further analysis.

This workflow was completed with the three feature groups defined above (the dataset with correlation values less than 0.7 removed, the PCA dataset, and the full dataset). The dataset containing all 161 items in the clinical inventories was the most robust with respect to ARI and bootstrapping analyses for the GW dataset (data not shown). Therefore, subsequent analyses to determine if external clinical and biological markers not used as features in the cluster analyses were associated with different symptom profile groups, were completed with these data, and are described below. After using the described process with the GW dataset, the GW and MIRECC dataset were merged on all available clinical features. As the MIRECC dataset only had 61 features (the full CAPS, a summary score for depression [HAM-D]), and 10 summary scores (from the SCL-90), a Gulf War Summary Scores dataset (GS) with 61 features was also created.

The cluster workflow described above was employed with the GS dataset (61 features), the MS dataset alone (61 features), and the combined GW and MIRECC dataset, referred to as the CS dataset (61 features).

Descriptive statistical analysis on the imaging, neuropsychiatric, and neuroendocrine measures was conducted using analysis of covariance (ANCOVA) to determine differences in biomarkers across clusters. Age and years of education were covariates in the ANCOVA for analysis of neuropsychiatric measures, age was a covariate in the evaluation of the imaging volumes, and age and salivary dexamethasone level were covariates in the analysis of the cortisol measures. As this was an exploratory study, all results with p-values of <0.05 were reported as statistically significant. None of the ANCOVAs using the genetic data were statistically significant at $p<.05$, so these are not reported. Post-hoc differences between clusters were identified using the Tukey-Kramer criteria.

To facilitate comparison between the results from the analyses using phenotypes derived from psychiatric symptoms to differences identified using the DSM PTSD phenotype, descriptive statistical analysis on the clinical and biological data was also conducted across all biomarkers using a male PTSD group vs. a male No PTSD group (e.g. the males in the datasets who did not meet criteria for a DSM diagnoses of PTSD). ANCOVA was used with age and years of education as covariates in the analysis of neuropsychiatric measures, age as a covariate in the evaluation of imaging volumes, and age and salivary dexamethasone level as covariates in the analysis of the cortisol measures. As with the cluster analyses, all results with p-values of <0.05 were reported as statistically significant. Post-hoc differences between clusters were identified using the Tukey-Kramer criteria.

Presentation of Results

Clusters were presented as symptom profiles with normalized mean values of each available symptom in each of the three inventories plotted on a line graph for each cluster identified in every analysis. All of the mean values were connected by a line delineating a symptom profile for each cluster. For ease of interpretation of results, clusters were numbered based on the level of their symptom profiles. The clusters with the overall highest (e.g. most severe) symptom is always delineated as cluster number 1, followed by those with lower symptom profiles. For example, for a 3 clusters solution, cluster 1 has the highest symptom profile, followed by cluster 2, and cluster 3 has the lowest symptom profile.

Additionally, while all results from all the above described analyses are included in this dissertation, the majority of the detailed findings have been moved to the appendices. In the body of the main, two analyses are highlighted that produced salient findings with regards to prior inconsistencies of biomarker correlates with the DSM PTSD phenotype in the literature, to demonstrate potential advantages to the approach used in these analyses. These two inconsistencies are that of reduced hippocampal size, and alterations in the Hypothalamic-Pituitary-Adrenal (HPA) Axis, in individuals with PTSD as compared with controls.

Finally, a rich clinical phenotype is introduced to allow for visualization of significant differences across multiple groups in one concise figure. In this paper, all biomarkers in a cluster that were significantly higher or lower than the mean value of biomarker in the entire sample were included in the rich clinical phenotype, and delineated as higher or lower than expected. For symptoms data, any symptom that significantly differed from its expected value with regards to the cluster's symptom profile was included.

Chapter 5: Cluster Analyses of Self-Report Symptoms in the Gulf War Dataset

In this chapter, the results of the cluster analyses in the Gulf War dataset (GW) are described. Initially, the results of the CH and wss indices are displayed as they were used to inform the number of clusters created using the HCA and k-means algorithms. ARI and bootstrapping measures determine the robustness of these solutions. Symptom profiles are presented along with tables and figures delineating the clinical and biological variables that differ across the groups of patients within the clusters.

As described in the methods section, the Gulf War dataset was clustered in two ways: 1) the GW full-feature dataset, which is described in this chapter, with the full set of psychiatric symptoms ascertained from the 161 itemized questions available; and 2) the GS data, which is located in Appendix D2, with the reduced set of 61 psychiatric symptoms, including summary measures, so that these data could ultimately be pooled with the MIRECC dataset. The GW dataset is comprised of 238 males, 33 with PTSD diagnoses, 26 with a history of MDD, and 60 with a lifetime history of alcohol dependence, as described in Table 5.1. The cluster descriptions for the GW k=2, k=3, k=4, and k=5 analyses are shown in Table 5.1

Table 5.1. Details of Cluster Solutions for k=2, k=3, k=4, and k=5

Dataset (# of Clusters) ARI Bootstrap	Cluster (# of Individuals in Cluster)	PTSD	MDD	Other DSM Avg (Range)	Total DSM Avg (Range)	ALC Dependence	$\chi 2$	CTR	$\chi 2$
GW (2)	Cluster2.1 (60)	27	17	0.8(0-4)	2.0(0-7)	24	9.3***	14	0.6
ARI: 0.24	Cluster2.2 (178)	6	8	0.6(0-7)	0.9(0-8)	36		33	
Boot: 0,0									
GW (3)	Cluster3.1 (14)	12	7	0.7(0-3)	2.9(0-6)	5	13.7***	3	0.1
ARI: 0.56	Cluster3.2 (70)	20	13	0.8(0-1)	2.4(0-7)	28		13	
Boot: 2,0,3	Cluster3.3 (154)	1	5	0.4(0-7)	1.2(0-15)	27		28	
GW (4)	Cluster4.1 (14)	12	7	0.7(0-3)	2.9(0-6)	5	20.3****	6	14.4***
ARI: 0.39	Cluster4.2 (54)	15	12	0.3(0-3)	1.7(0-7)	23		9	
Boot: 25,17,1,17	Cluster4.3 (91)	6	4	0.4(0-6)	1.1(0-8)	23		25	
	Cluster4.4 (79)	0	2	0.9(0-2)	0.5(0-3)	9		7	
GW (5)	Cluster5.1 (14)	12	7	0.7(0-3)	2.9(0-6)	5	35.9****	6	17.4***
	Cluster5.2 (39)	19	6	0.4(0-3)	1.9(0-9)	18		3	
ARI: 0.74	Cluster5.3 (26)	0	7	0.2(0-3)	1.2(0-4)	6		9	
Boot:1,20,1,0,0	Cluster5.4 (86)	2	3	0.4(0-6)	1.1(0-8)	22		25	
	Cluster5.5 (73)	0	2	0.1(0-2)	0.6(0-3)	9		5	

GW: Gulf War Sample
ARI: Adjusted Rand Index between the k-means cluster and hierarchical cluster solutions. Boot shows the number of times each cluster dissolved during 100 bootstraps of the cluster analyses.
CTR: Childhood Trauma
*** p<.001; **** p<.0001

The Calinski-Harabasz (CH) Index and Within Sum of Squares (wss) results are shown in Figure 5.1. The optimal number of clusters was k=2, based on the CH index. In contrast, the wss analysis was at a minimum between k=4 and k=6. The Adjusted Rand Index (ARI) favored cluster solutions where k=5 (0.74), followed by those solutions where k=3 (0.56). When

bootstrap was used with the k-means algorithm, solutions with k=2, k=3, and k=5 had clusters that were stable at least 80 percent of the time.

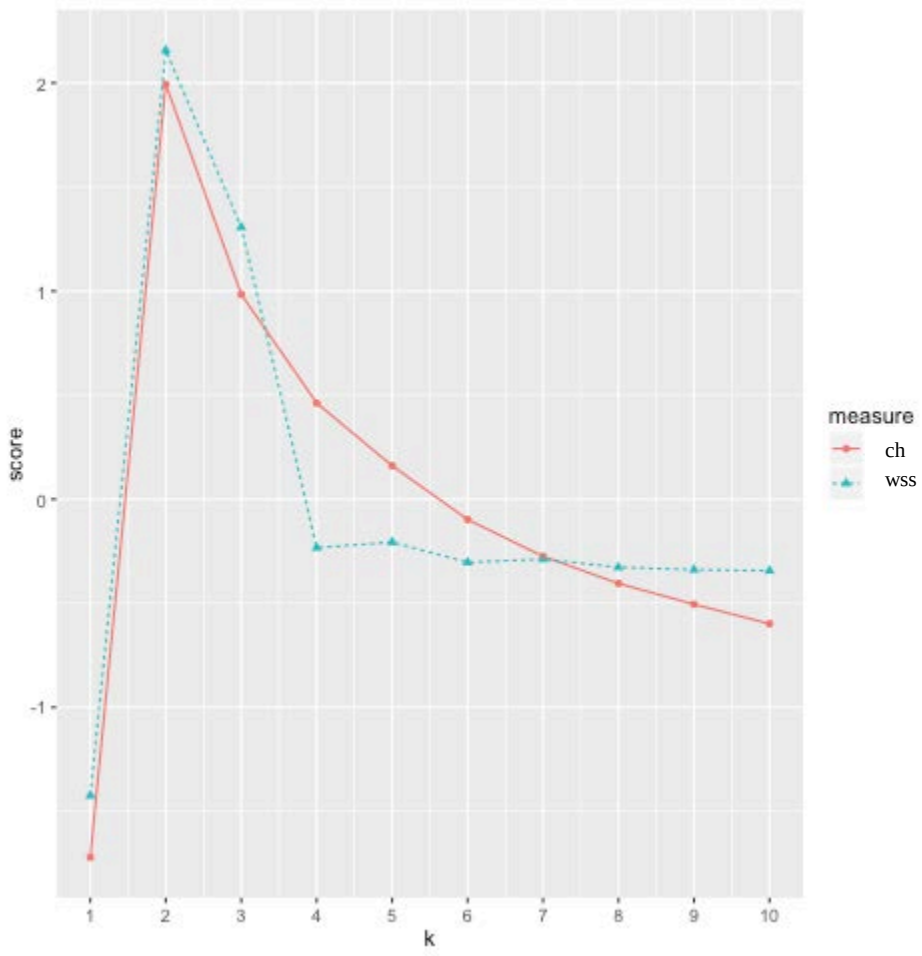

Figure 5.1. Details of Cluster Solutions for k=2, k=3, k=4, and k=5

As can be seen in Table 5.1, DSM diagnoses do not coincide with cluster delineation. Across all cluster solutions, the individuals with a diagnosis or PTSD were split across at least two clusters. This stratification across clusters for DSM diagnosis also occurred with a diagnosis of Major Depressive Disorder, and Alcohol Dependence, demonstrating that the clusters derived through itemized psychiatric symptoms do not converge with these DSM diagnoses.

The k=5 cluster solution was the most robust solution according to the ARI (0.74) and the bootstrap evaluation, where four of the five clusters were stable for >98 percent of perturbations and the fifth cluster was stable for 80 percent of perturbations (Table 5.1). Cluster 5.1 included a substantial number of individuals (12/14) with PTSD, as well as 7/14 with MDD, and 5/14 with alcohol dependence. Cluster 5.2 comprised all but two of the remaining individuals with PTSD (19/33), 6/25 individuals with MDD, and 18/69 patients with alcohol dependence. However, symptom graphs (Figures 5.2a–5.2c) show that Cluster 5.2 (DSM avg=1.9) has a symptom profile that appears very similar to that of Cluster 5.3 (DSM avg=1.2), despite the fact that Cluster 5.3 does not include any individuals with a diagnosis of PTSD. The incidence of alcohol dependence across all five clusters differed significantly ($\chi 2=35.9$, $p<.0001$; Table 5.1), with the percentage of individuals with alcohol dependence being the greatest in 5.1 and least in 5.5, generally decreasing as the symptom profiles lowered. This decreasing pattern as the cluster symptom profile decreased was also generally seen in the incidence of childhood trauma ($\chi 2=17.4$, $p<.001$; Table 5.1).

Table 5.2 summarizes the neuropsychiatric performances for each cluster in the k=5 cluster solution. Individuals in Cluster 5.1 performed worse in the Executive Functioning ($p<.05$), Performance IQ ($p<.05$, $p<.01$), and Verbal IQ ($p<.05$, $p<.01$, $p<.001$) domains than those in Clusters 5.4 and 5.5 (Table 5.5, Figure 5.5j). Cluster 5.3 also had Performance IQ scores

that were worse than clusters 5.4 and 5.5 ($p<.05$, $p<.01$). Interestingly, Cluster 5.1 also performed more poorly than individuals in Cluster 5.2 (the cluster comprising the remaining individuals with PTSD) on one measure of Verbal IQ ($p<.001$) (Table 5.2, Figure 5.3a). Cluster 5.1 has larger average right lenticular white matter volumes than Clusters 5.3 and 5.4 ($p<0.05$) (Table 5.2, Figure 5.3b). Cluster 5.1 also has the lowest average right frontal and parietal CSF volumes ($p<0.05$; Table 5.2, Figure 5.3b). Additionally, Cluster 5.2 has smaller average right hippocampal volumes than Cluster 5.3 ($p<0.05$; Table 5.2, Figure 5.3b). There are no significant differences between the groups in cortisol measures or genetic measures.

In an attempt to integrate all of these multimodal markers into a concise graphical display, a Rich Clinical Phenotype is shown in Figure 5.4. In this figure, all five clusters are represented as designated in the legend. This figure allows the comparison across all five clusters of differences in clinical and biological markers. The self-report symptoms that are displayed in the figure as respectively colored bars are those where a cluster had a value of a symptom that was significantly higher or lower than the expected value, as determined by the symptom profile mapping. For instance, although Cluster 5.2 generally had a higher level of symptoms than Cluster 5.3, as shown in Figures 5.2a–c, there is a group of symptoms for which 5.3 has significantly higher values than 5.2. In contrast, the biological markers shown in this figure represent a significant difference in the value of these markers compared with the means of the other clusters.

Looking closely at Figure 5.4 it can be seen that Clusters 5.2 (blue) and 5.3 (purple) differ with respect to 5.2 having an array of worse than expected symptoms when compared with Cluster 5.3, as well as a significantly smaller hippocampal size than individuals in Cluster 5.3.

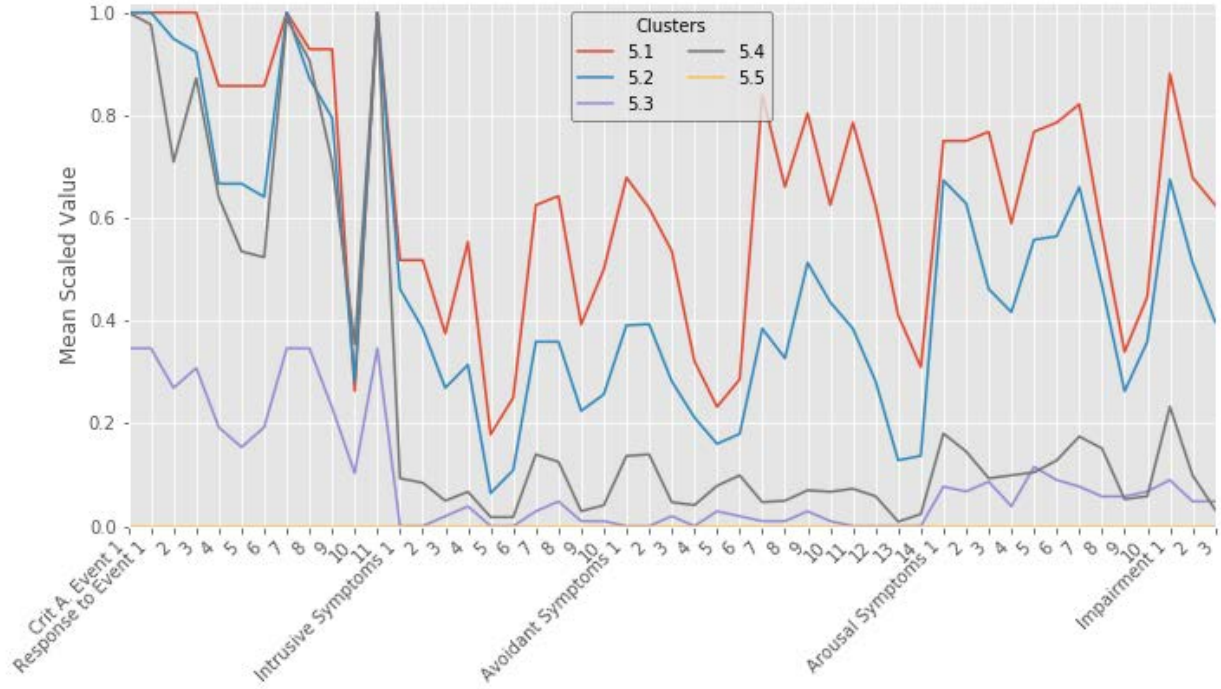

Figure 5.2a. Symptom Profiles for All Items in the Clinician Assessment for PTSD in GW Dataset (k=5)

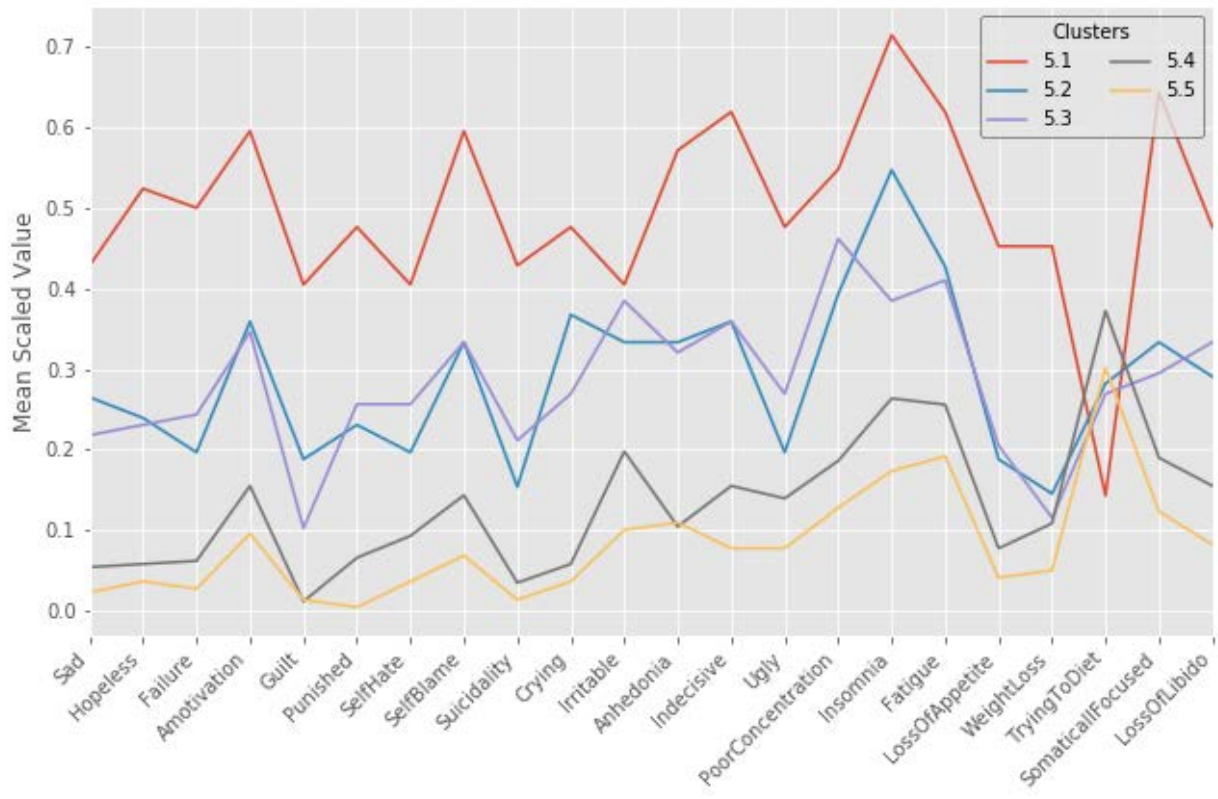

Figure 5.2b. Symptom Profiles for All Items in the Beck Depression Inventory in GW Dataset (k=5)

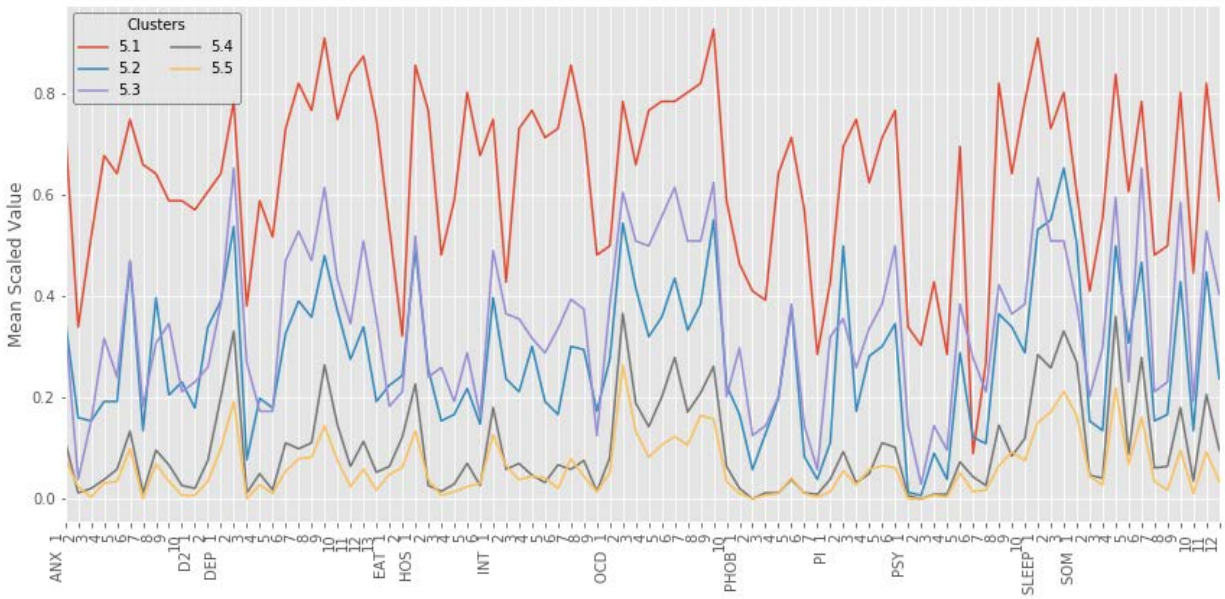

Figure 5.2c. Symptom Profiles for All Items in the Symptom Checklist-90 in GW Dataset (k=5)

Test	Cluster 1 Mean	S.D.	Cluster 5.9 Mean	S.D.	Cluster 5.2 Mean	S.D.	Cluster 4 Mean	S.D.	Cluster 5 Mean	S.D.	ANCOV A P-value	Sig	F	Post Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=4,225)														
Executive Functioning 1	8.29E+00	2.86E+00	9.61E+00	9.41E+00	1.09E+01	9.42E+00	1.11E+01	9.05E+00	1.07E+01	9.05E+00	8.16E-05	**	9.52	4>1*
Executive Functioning 2	9.92E+00	2.22E+00	1.06E+01	9.19E+00	1.18E+01	2.90E+00	1.29E+01	2.40E+00	1.17E+01	2.95E+00	4.38E-05	**	9.9	4>1*
Intrusions 1	4.85E+00	4.97E+00	6.44E+00	5.95E+00	9.55E+00	9.90E+00	7.54E+00	6.85E+00	4.55E+00	4.45E+00	9.98E-02	*	2.56	NS.
Intrusions 2	9.06E+00	9.24E+00	9.88E+00	9.97E+00	1.82E+00	1.95E+00	4.54E+00	4.97E+00	2.85E+00	4.45E+00	9.92E-02	*	2.56	NS.
Memory 1	8.92E+00	2.96E+00	9.74E+00	4.05E+00	1.12E+01	9.45E+00	1.15E+01	9.04E+00	1.12E+01	9.15E+00	1.12E-02	*	9.54	
Memory 2	1.24E+01	2.70E+00	1.17E+01	2.94E+00	1.09E+01	9.68E+00	1.00E+01	2.91E+00	1.20E+01	2.86E+00	1.85E-02	*	9.04	NS.
Performance IQ 1	1.24E+01	2.70E+00	1.17E+01	2.94E+00	1.09E+01	9.68E+00	1.00E+01	2.91E+00	1.20E+01	2.86E+00	1.85E-02	***	4.87	4>9*,5>9**
Performance IQ 2	6.06E+01	1.95E+01	6.85E+01	1.28E+01	5.91E+01	1.92E+01	6.28E+01	1.65E+01	7.14E+01	1.95E+01	8.70E-04	**	4.15	5>1*
Verbal IQ 1	9.48E+01	1.10E+01	9.97E+01	1.91E+01	9.91E+01	1.18E+01	4.27E+01	1.18E+01	4.56E+01	1.14E+01	2.89E-05	**	9.76	4>1*,5>1*
Verbal IQ 9	9.81E+01	1.16E+01	4.72E+01	8.08E+00	4.18E+01	1.17E+01	4.69E+01	8.81E+00	4.66E+01	1.05E+01	5.52E-05	***	5.72	4>1**,2>1**, 5>1**
Imaging Volumes (D.F.=4,88)														
Right Hippocampal	2.07E-03	2.00E-04	2.01E-03	9.00E-04	2.26E-03	9.00E-04	2.14E-03	9.00E-04	2.12E-03	2.00E-04	2.66E-02	*	2.83	9>2*
Right Lenticular White Matter	2.71E-03	8.19E-04	2.29E-03	4.41E-04	2.99E-03	4.62E-04	2.25E-03	4.65E-04	2.19E-03	4.34E-04	2.96E-02	*	2.76	1>4*,1>5*
Cerebellar White Matter	2.76E-02	2.88E-03	2.61E-02	9.51E-03	2.88E-02	6.90E-03	2.55E-02	4.78E-03	2.59E-02	4.59E-03	9.40E-02	*	2.67	2>1*
Right Parietal CSF	1.91E-02	4.90E-03	2.18E-02	4.70E-03	2.26E-02	4.70E-03	2.20E-02	5.00E-03	2.40E-02	5.20E-03	1.45E-02	*	9.2	5>1*
Left Parietal CSF	1.85E-02	4.90E-03	2.17E-02	4.60E-03	2.15E-02	4.80E-03	2.15E-02	4.40E-03	2.91E-02	5.50E-03	1.92E-02	*	9.03	5>1*
Left Occipital CSF	4.27E-03	1.20E-03	5.26E-03	1.20E-03	4.85E-03	1.20E-03	4.74E-03	1.20E-03	5.70E-03	1.50E-03	1.32E-04	***	6.11	5>1**
Left Temporal CSF	1.95E-02	9.15E-03	1.40E-02	2.59E-03	1.92E-02	2.99E-03	1.25E-02	2.92E-03	1.45E-02	9.24E-03	9.01E-02	*	2.75	NS.
Subcortical CSF	8.95E-04	2.34E-04	1.05E-03	2.71E-04	1.05E-03	1.97E-04	1.05E-03	1.97E-04	1.08E-03	2.44E-04	2.58E-02	*	2.85	4>1*

D.F., degrees of freedom

S.D., standard deviation

N.S., not significant

TK, Tukey-Kramer

*, p<.05; **, p<.01; ***, p<.001

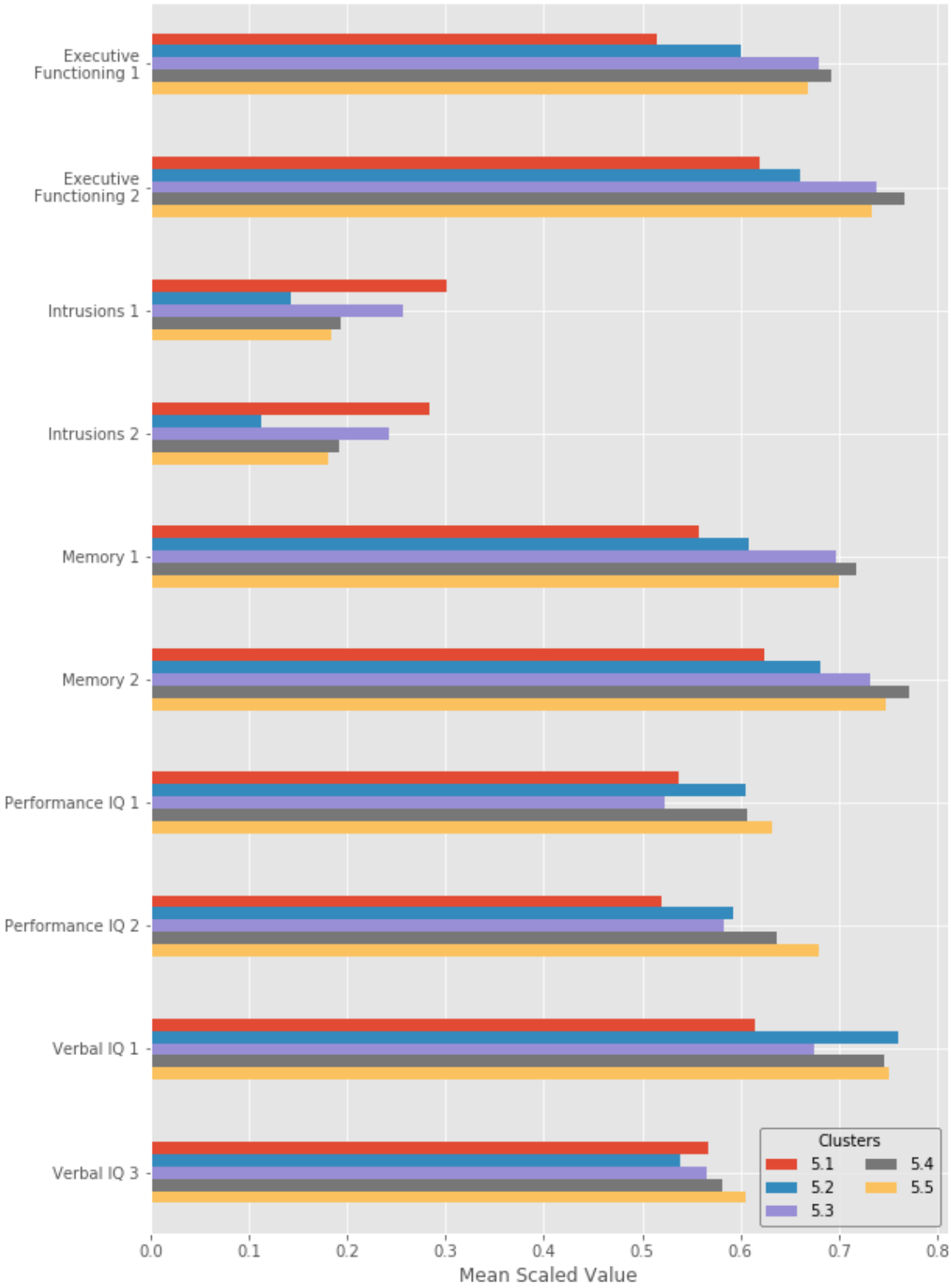

Figure 5.3a. Neuropsychiatric Markers with Significant Differences Across Clusters in GW Dataset (k=5)

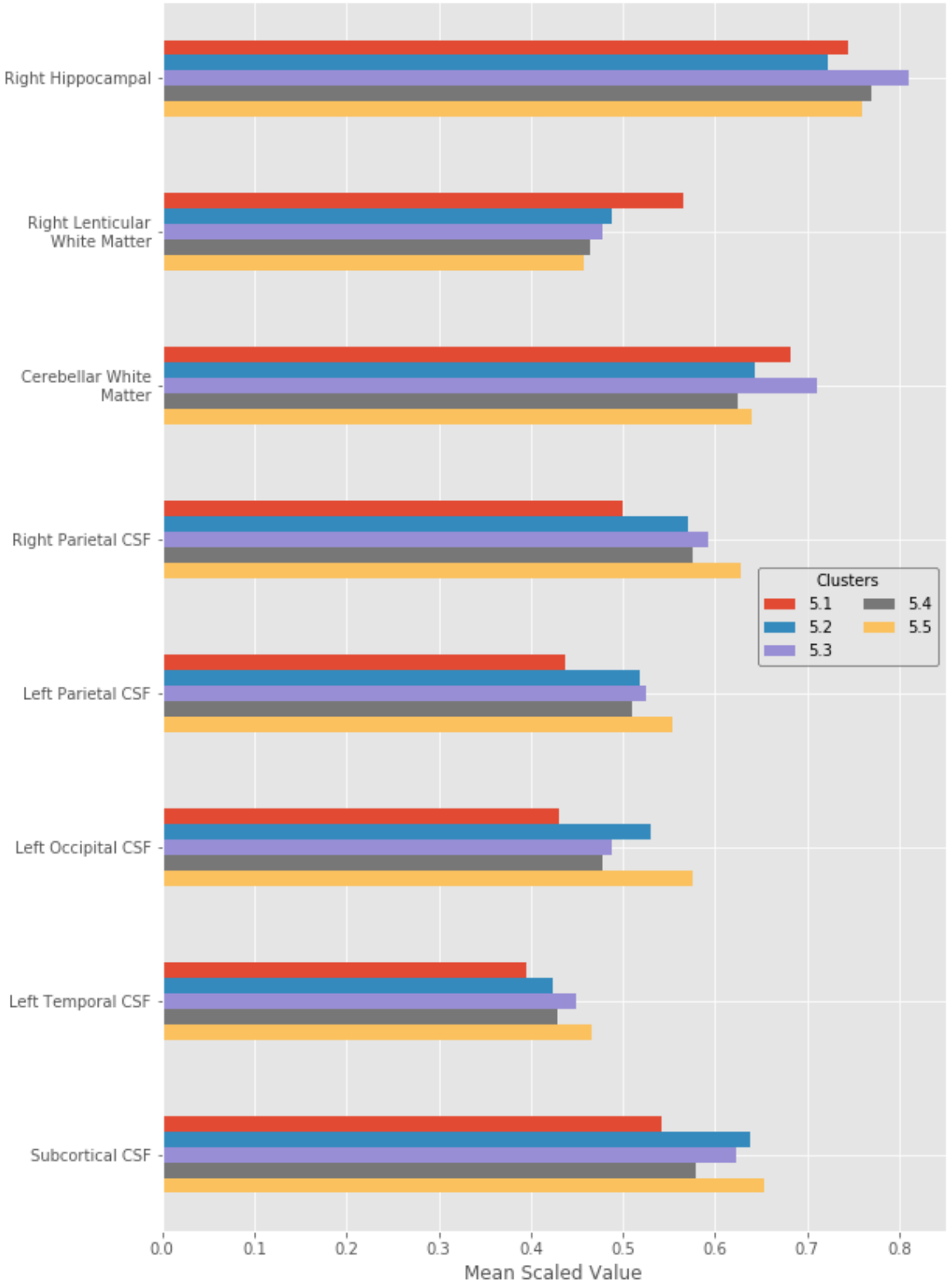

Figure 5.3b. Imaging Markers with Significant Differences Across Clusters in GW Dataset (k=5)

6a shows

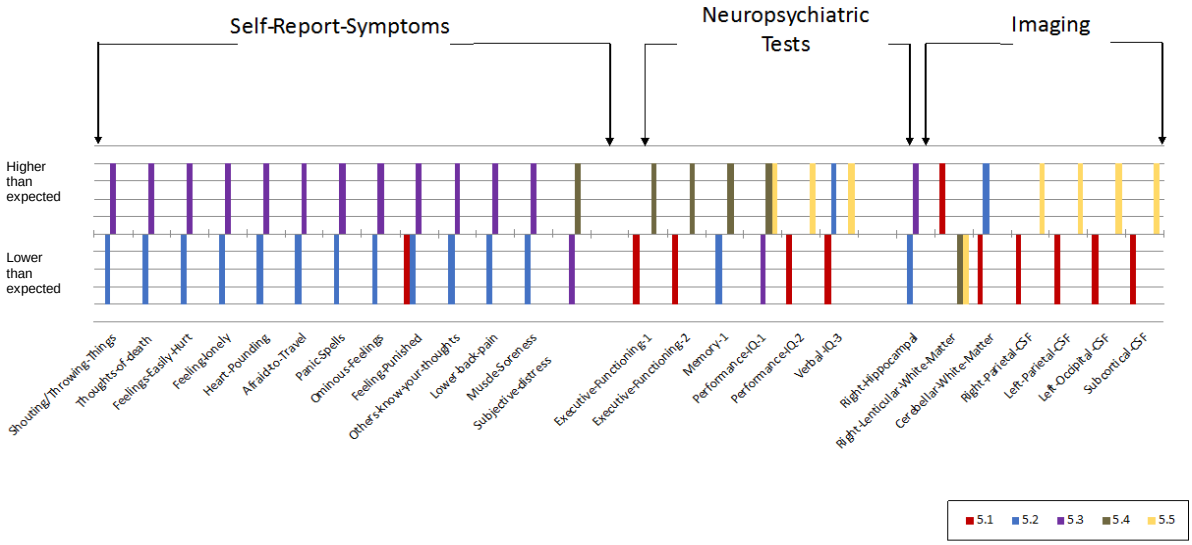

Figure 5.4. Rich Clinical Phenotype of GW k=5 Solution

To contrast the differences between using highly granular psychiatric symptom data vs. DSM PTSD diagnosis, individuals with a diagnosis of PTSD in the GW dataset are compared with those without the diagnoses below. Figures 5.5a–c show the different symptom profiles of these two groups. Unlike the previous five-cluster solution, in these clinical inventories, the symptom profiles only cross at all at one item. The questions asked if the subject was trying to diet, and the responses indicated that individuals who reported a higher level of distressing psychiatric symptoms were not as likely to be trying to diet as those with less distress. This inverse association with respect to all of the other symptoms mirrored findings in the cluster solution, as can be seen in figure 5.2b. Aside from this single exception, the PTSD group always has more distress, and hence higher levels across all symptoms in all three clinical inventories.

Similarly, Figure 5.6a shows that in all of the neuropsychiatric tests that were significantly different between groups, the PTSD group had lower or “worse” scores on the

neuropsychiatric tests (the statistical significance of biomarker differences between the PTSD group and the No PTSD group are reported in Table 5.3). Figure 5.6b shows far fewer significant differences than the five cluster solutions, with right lenticular and right caudate white matter volumes being greater in the PTSD group than in the unaffected group, and the left temporal CSF values being less in the PTSD group than the unaffected group. Implications of this finding are elaborated on in the discussion section.

Finally, the rich clinical phenotype in Figure 5.7 of these two groups represents all of these findings. When compared with the rich clinical phenotype for the five cluster solutions in Figure 5.4, it can be seen that the five clusters have significantly more clinical and biological heterogeneity than the PTSD and No PTSD groups.

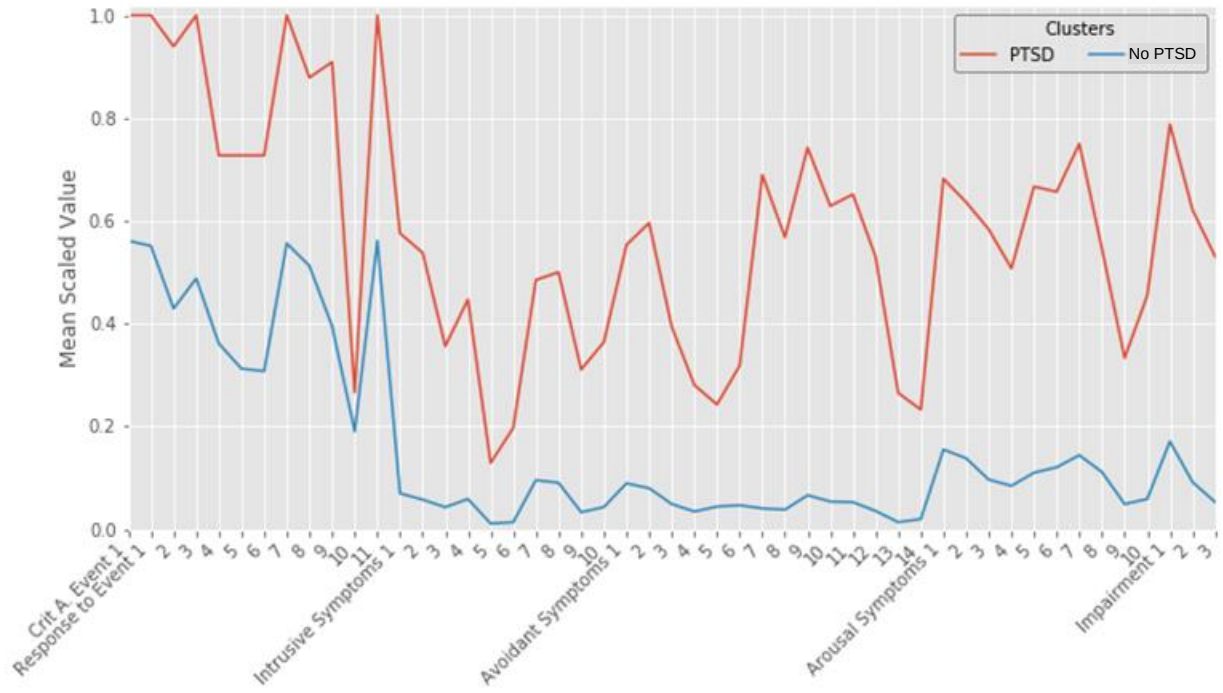

Figure 5.5a. Symptom Profiles for All Items in the Clinician Assessment for PTSD vs. No PTSD in GW Dataset

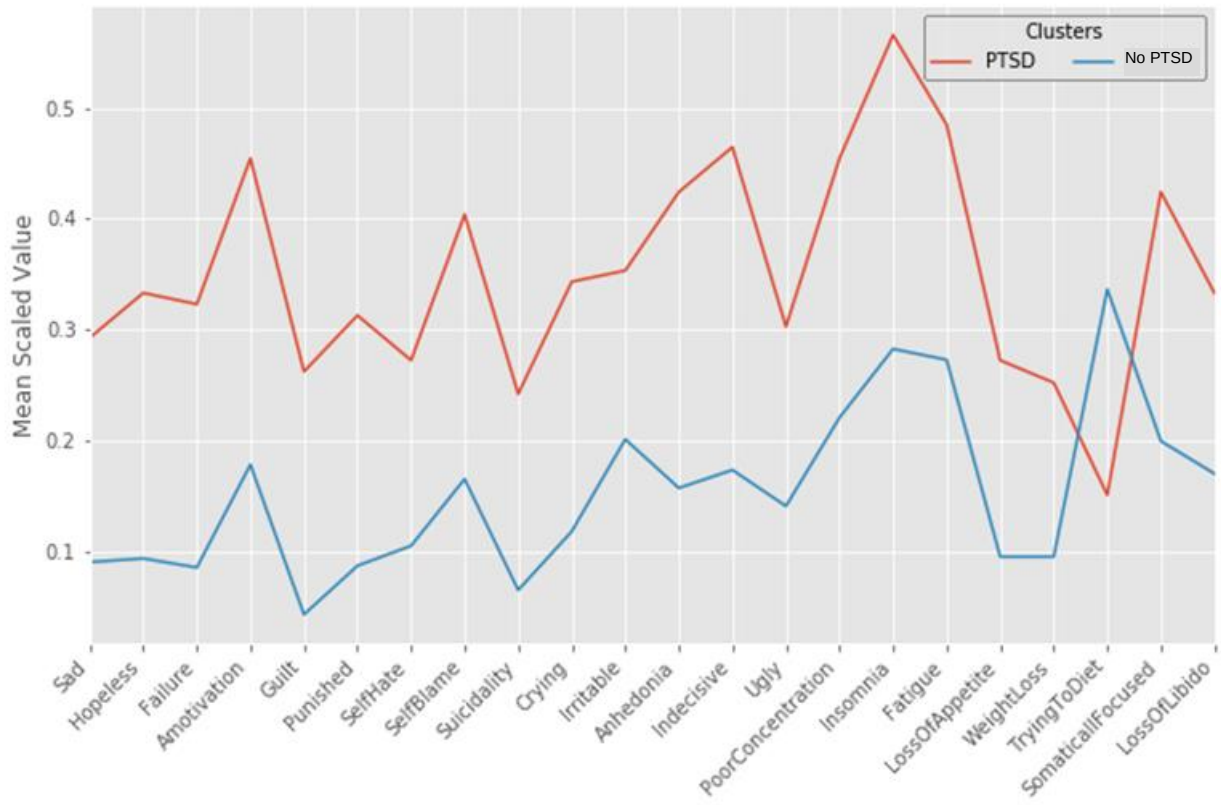

Figure 5.5b. Symptom Profiles for All Items in the Beck Depression Inventory in GW Dataset for PTSD vs. No PTSD

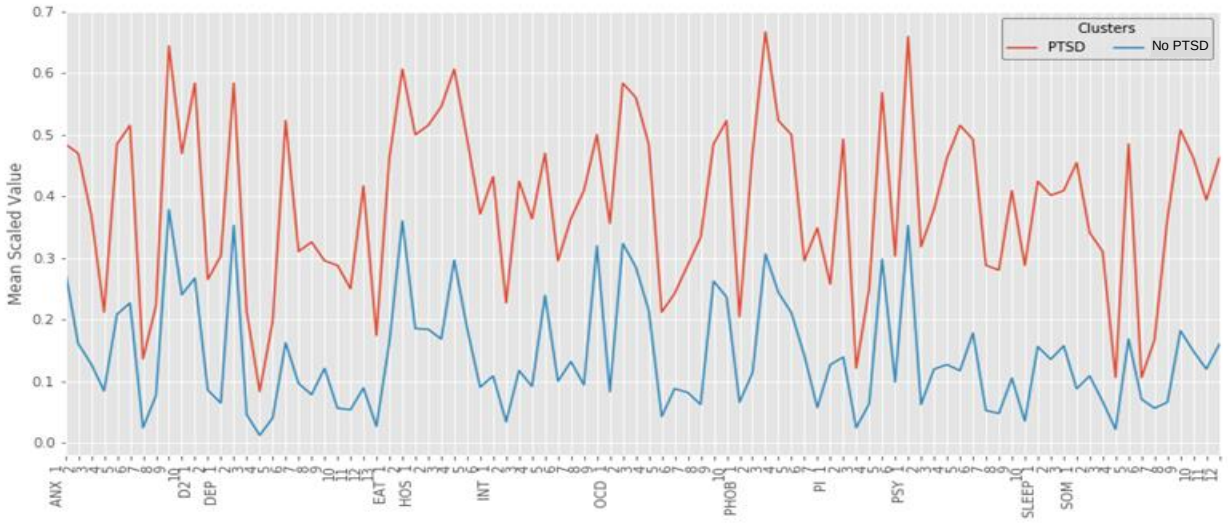

Figure 5.5c. Symptom Profiles for All Items in the Symptom Checklist-90 Inventory in GW Dataset for PTSD vs. No PTSD SCL-90

Table 5.3. Biomarker Differences Between PTSD vs. No PTSD in the GW

Test	PTSD (Clus.1) Mean	S.D.	No PTSD (Clus.2) Mean	S.D.	ANCOVA P-value	Sig	F	Post-Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=1,228)								
Learning 1	5.47E+00	1.83E+00	6.24E+00	1.86E+00	2.76E-02	*	4.92	2>1*
Learning 2	1.13E+01	2.76E+00	1.24E+01	2.60E+00	2.92E-02	*	4.82	2>1*
Learning 3	4.58E+01	1.04E+01	5.11E+01	9.84E+00	4.60E-03	**	8.19	2>1**
Executive Functioning 1	9.13E+00	3.16E+00	1.08E+01	3.14E+00	5.56E-03	**	7.84	2>1**
Executive Functioning 2	1.03E+01	3.01E+00	1.19E+01	2.73E+00	2.03E-03	**	9.75	2>1**
Memory 1	9.25E+00	3.43E+00	1.12E+01	3.25E+00	1.70E-03	**	10.09	2>1**
Memory 2	1.01E+01	3.35E+00	1.21E+01	2.84E+00	4.53E-04	***	12.67	2>1***
Performance IQ 2	3.73E+01	1.23E+01	4.30E+01	1.19E+01	8.25E-03	**	7.10	2>1**
Verbal IQ 4	1.81E+01	4.34E+00	1.96E+01	4.19E+00	3.76E-02	*	4.38	2>1*
Imaging Volumes (D.F.=165)								
Right Caudate White Matter	9.12E-04	5.11E-04	7.23E-04	2.33E-04	2.09E-03	**	9.78	1>2**
Right Lenticular White Matter	2.45E-03	6.30E-04	2.23E-03	4.48E-04	2.97E-02	*	4.81	1>2*
Left Temporal CSF	1.26E-02	2.33E-03	1.39E-02	3.08E-03	1.57E-02	*	5.96	1<2*

D.F.: degrees of freedom
S.D.: standard deviation
N.S.: not significant;
TK: Tukey-Kramer
PTSD is Cluster 1; No PTSD is Cluster 2
*, p<.05; **, p<.01; ***, p<.001

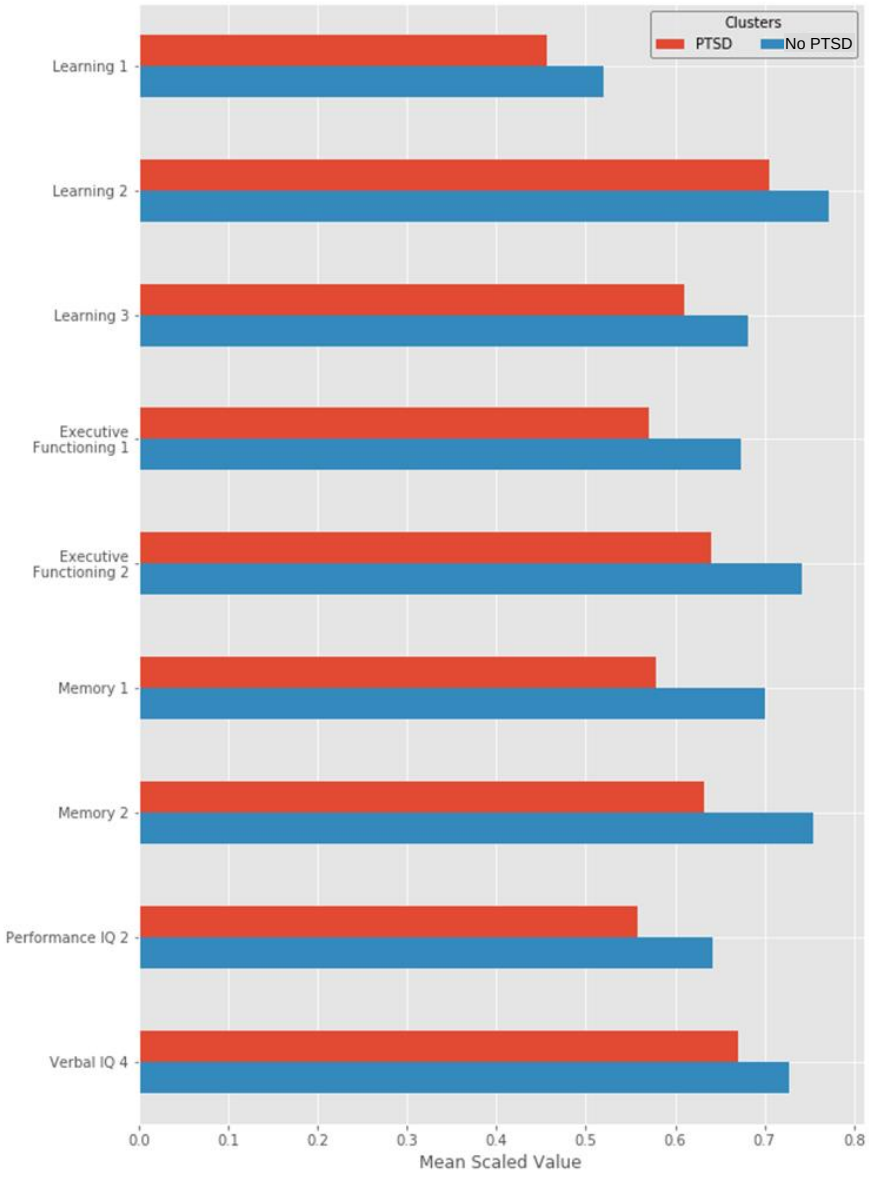

Figure 5.6a. Neuropsychiatric Markers with Significant Differences in PTSD vs No PTSD in GW Dataset

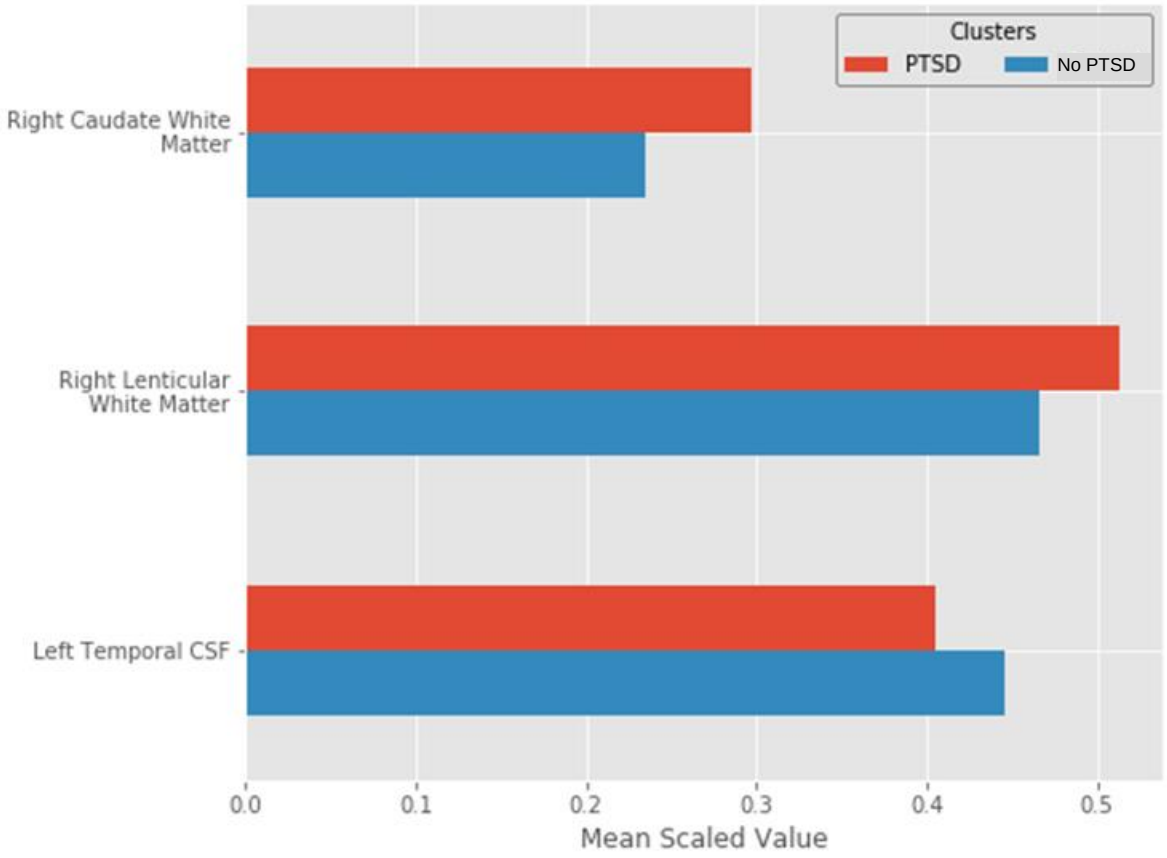

Figure 5.6b. Imaging Markers with Significant Differences in PTSD vs. No PTSD in GW Dataset

Table 5.3. Biomarker Differences Between PTSD vs. No PTSD in the GW

Test	PTSD (Clus.1) Mean	S.D.	No PTSD (Clus.2) Mean	S.D.	ANCOVA P-value	Sig	F	Post-Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=1,228)								
Learning 1	5.47E+00	1.83E+00	6.24E+00	1.86E+00	2.76E-02	*	4.92	2>1*
Learning 2	1.13E+01	2.76E+00	1.24E+01	2.60E+00	2.92E-02	*	4.82	2>1*
Learning 3	4.58E+01	1.04E+01	5.11E+01	9.84E+00	4.60E-03	**	8.19	2>1**
Executive Functioning 1	9.13E+00	3.16E+00	1.08E+01	3.14E+00	5.56E-03	**	7.84	2>1**
Executive Functioning 2	1.03E+01	3.01E+00	1.19E+01	2.73E+00	2.03E-03	**	9.75	2>1**
Memory 1	9.25E+00	3.43E+00	1.12E+01	3.25E+00	1.70E-03	**	10.09	2>1**
Memory 2	1.01E+01	3.35E+00	1.21E+01	2.84E+00	4.53E-04	***	12.67	2>1***
Performance IQ 2	3.73E+01	1.23E+01	4.30E+01	1.19E+01	8.25E-03	**	7.10	2>1**
Verbal IQ 4	1.81E+01	4.34E+00	1.96E+01	4.19E+00	3.76E-02	*	4.38	2>1*
Imaging Volumes (D.F.=165)								
Right Caudate White Matter	9.12E-04	5.11E-04	7.23E-04	2.33E-04	2.09E-03	**	9.78	1>2**
Right Lenticular White Matter	2.45E-03	6.30E-04	2.23E-03	4.48E-04	2.97E-02	*	4.81	1>2*
Left Temporal CSF	1.26E-02	2.33E-03	1.39E-02	3.08E-03	1.57E-02	*	5.96	1<2*

D.F.: degrees of freedom
S.D.: standard deviation
N.S.: not significant;
TK: Tukey-Kramer
PTSD is Cluster 1; No PTSD is Cluster 2
*, p<.05; **, p<.01; ***, p<.001

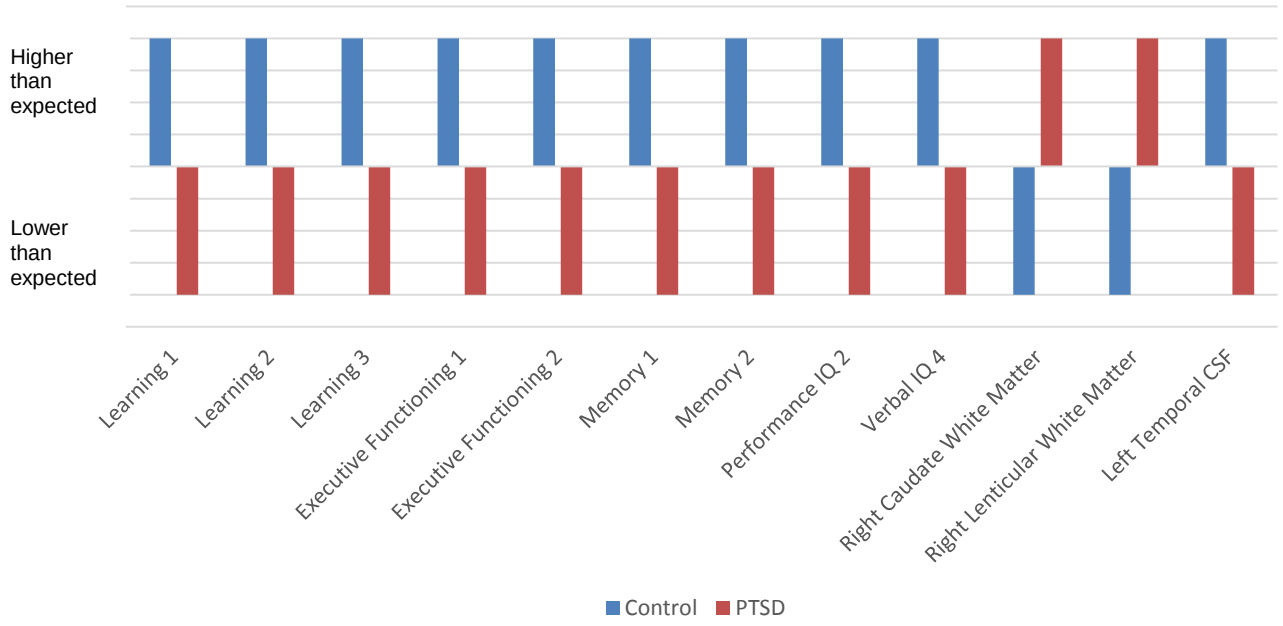

Figure 5.7. Rich Clinical Phenotype of PTSD vs. No PTSD in GW Dataset

Chapter 6: Cluster Analyses of Summary Self-Report Symptoms in Combined Gulf War and MIRECC Datasets

To examine the relationship between self-report symptom profiles and clinical and biological markers in the entire dataset, the Gulf War and MIRECC datasets were merged to form the Combined Summary (CS) dataset. The total number of features used for the CS analysis was 61 (as opposed to 161 in the GW), with all 50 items from the CAPS, 10 summary scores from the SCL-90, and one summary item from the HAM-D. There were 309 individuals in this dataset, 77 with PTSD, 41 with MDD, and 95 with alcohol dependence.

Table 6.1. Details of Cluster Solutions for k=2, k=3, k=4, and k=5

Dataset (# of Clusters) ARI Bootstrap	Cluster (# of Individuals in Cluster)	PTSD	MDD	Other DSM Avg (Range)	Total DSM Avg (Range)	ALC Depen-dence	$\chi 2$	Child-hood Trauma	$\chi 2$
CS (2)	Cluster2.1 (94)	75	28	1.4(0-8)	3(0-11)	44	13.5***	31	9.5**
ARI: 0.81	Cluster2.2 (215)	2	13	0.7(0-9)	1(0-9)	51		37	
Boot: 0,0									
CS (3)	Cluster3.1 (88)	75	27	1.4(0-8)	3(0-11)	40	21.6****	31	19.4****
ARI: 0.8	Cluster3.2 (132)	2	7	0.9(0-9)	1.3(0-9)	43		30	
Boot: 0,0,0	Cluster3.3 (89)	0	7	0.5(0-3)	0.7(0-4)	12		7	
CS (4)	Cluster4.1 (30)	29	17	1.4(0-5)	3.3(1-7)	11	21.1****	13	19.5***
ARI: 0.87	Cluster4.2 (69)	47	11	1.3(0-8)	2.7(0-11)	34		18	
Boot: 1,0,0,0	Cluster4.3 (120)	1	6	0.9(0-9)	1.3(0-9)	38		30	
	Cluster4.4 (89)	0	7	0.5(0-3)	0.7(0-4)	12		7	
CS (5)	Cluster5.1 (24)	24	14	1.3(0-4)	3.2(1-6)	8	19.9***	10	8.4**
ARI: 0.76	Cluster5.2 (42)	37	8	1.4(0-4)	2.9(0-11)	18		13	
Boot: 0,0,14,22,35	Cluster5.3 (46)	16	7	1.2(0-8)	2.2(0-10)	23		11	
	Cluster5.4 (108)	0	5	1(0-9)	1.3(0-9)	34		27	
	Cluster5.5 (89)	0	7	0.5(0-3)	0.7(0-4)	12		7	

CS: Combined Summary
ARI: Adjusted Rand Index between the k-means cluster and hierarchical cluster solutions.
BS: Bootstrap shows the number of times each cluster dissolved during 100 bootstraps of the cluster analyses.
**, p<.01; ***, p<.001; ****, p<.0001

The results of CH and wss analyses are shown in Figure 6.1. The k=2, k=3, and k=4 solutions all met the three criteria identified for cluster stability; the numbers of clusters were consistent with those in the CH and wss graphs, which had relatively high ARIs (all ≥ 0.8) and were stable >98 percent of the time in bootstrap analysis.

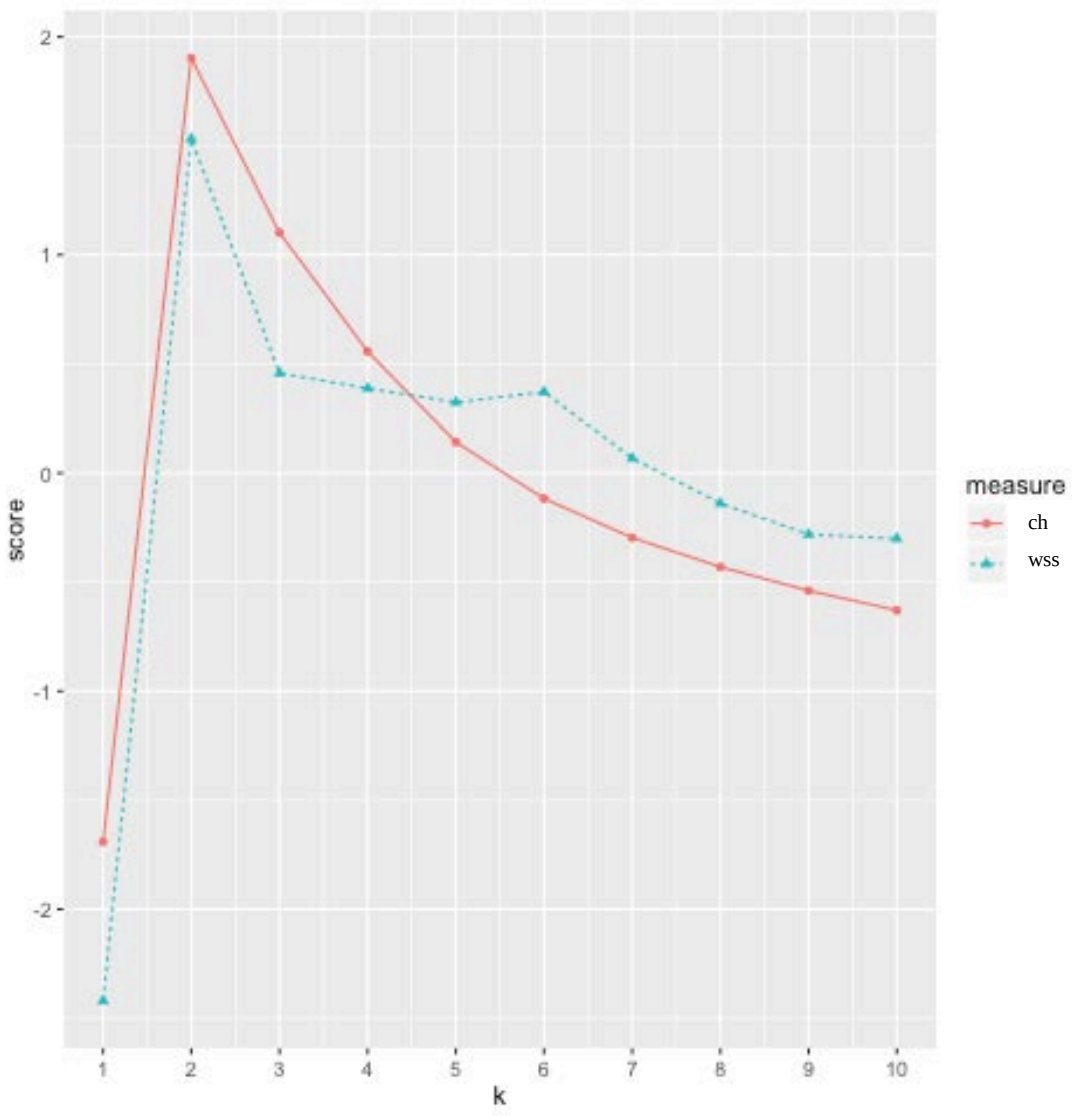

Figure 6.1. Details of Cluster Solutions for k=2, k=3, k=4, and k=5.

The four-cluster solution best met the criteria of stability, as shown as a minimum in the wss analysis and with the highest ARI, 0.87. Each of the clusters in this solution also had at least 99 percent stability, as revealed by bootstrapping. Of the four clusters in this solution, two comprised 76 of the 77 individuals with PTSD (Clusters 4.2 and 4.1). Cluster 4.1 had 38 percent (29/77) of the total individuals with PTSD, and these individuals made up 97 percent (29/30) of the entire cluster. The mean number of DSM diagnoses per individual in this cluster was 3.3, followed by Cluster 4.2 (average 2.7), Cluster 4.3 (average 1.3), and Cluster 4.4 (0.7). Cluster 4.2 had 61 percent of the individuals with PTSD (47/77), and these individuals made up 68 percent of this cluster. The average intensity of measured symptoms was higher in Cluster 4.1 than in those in the other three clusters, followed by Clusters 4.2, 4.3, and 4.4 (Figures 6.2a–b). The incidence of alcohol dependence across all four clusters differed significantly ($\chi 2=26.2$, $p<.00001$), with the percentage of individuals with alcohol dependence being the greatest in Cluster 4.2, followed by Clusters 4.1, 4.3, and 4.4 (Table 6.1). The incidence of childhood trauma also significantly differed across the four clusters, with the highest incidence in Cluster 4.1, followed by 4.2, 4.3, and 4. 4 ($\chi 2=17.5$, $p<.001$).

On average, individuals in Cluster 4.1 performed worse in all tested neuropsychiatric domains than individuals in the other clusters, including having poorer scores in Executive Function, Learning, Memory, and IQ domains than individuals in Cluster 4.2 and Cluster 4.3 ($p<.05$, $p<.01$, $p<.001$, $p<.0001$;Table 6.1, Figure 6.3a). These individuals also had larger average bilateral occipital and temporal white matter volumes, as well as larger average right temporal and left parietal white matter volumes, than those in Clusters 4.3 or 4.4 ($p<.05$, $p<.01$; Table 6.2, Figure 6.3b). Overall, CSF volumes were larger in Cluster 4.4 than in Clusters 4.1 or 4.2 ($p<.05$, $p<.01$, $p<.001$, $p<.0001$; Table 6.2, Figure 6.3b). Furthermore, individuals in Cluster

4.1 had lower basal cortisol levels than those in Cluster 4.4 ($p<.05$) and hypo-suppression of cortisol on Day 2 of the DST($p<.05$, $p<.001$; Table 6.2, Figure 6.3c). Figure 6.4 shows the Rich Clinical Phenotype of the four-cluster solution with visualization of the differences across all four clusters. While the p-values are unadjusted, and therefore may not be replicated in larger samples, this figure suggests a population of essentially two subgroups of individuals with PTSD and a fairly high level of other DSM disorders, and two subgroups of individuals without PTSD and with lower levels of psychiatric comorbidity. Both Clusters 4.1 and 4.2 fared worse on multiple neuropsychiatric tests than Clusters 4.3 and 4.4, but only Cluster 4.2 shows significantly greater white matter cortical volumes and cortisol suppression post DST, and only Cluster 4.1 shows significantly lower baseline AM cortisol levels.

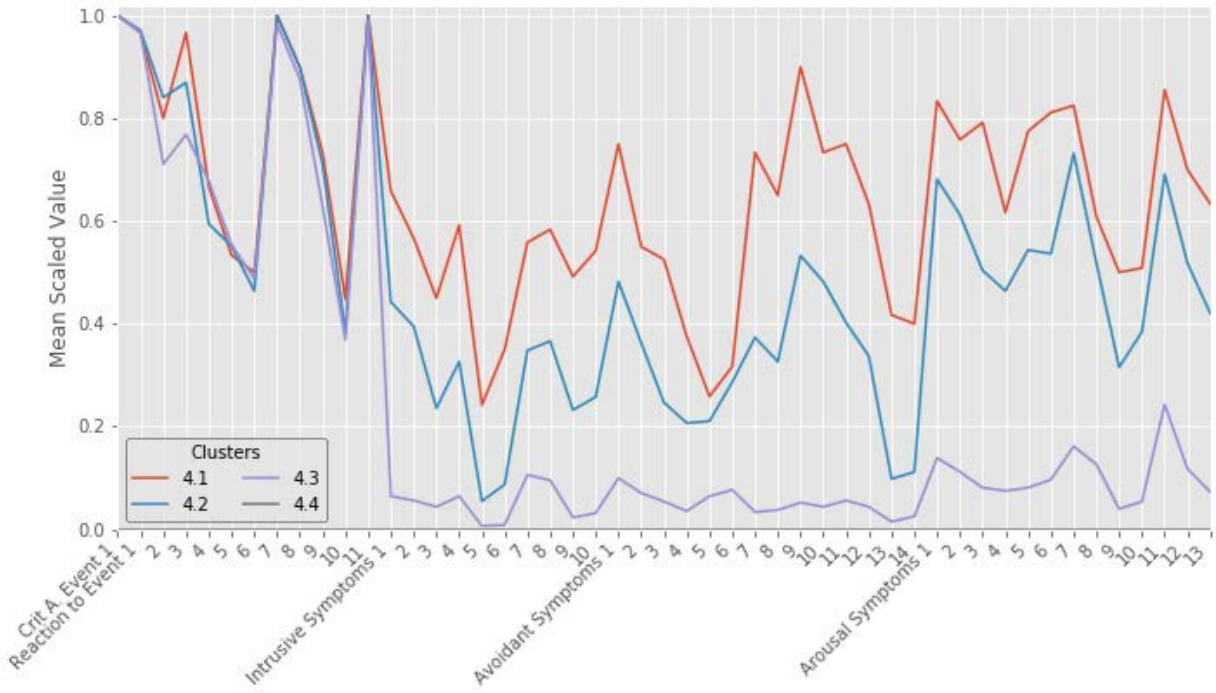

Figure 6.2a. Symptom Profiles for All Items in the Clinician Assessment for PTSD in the CS Dataset (k=4)

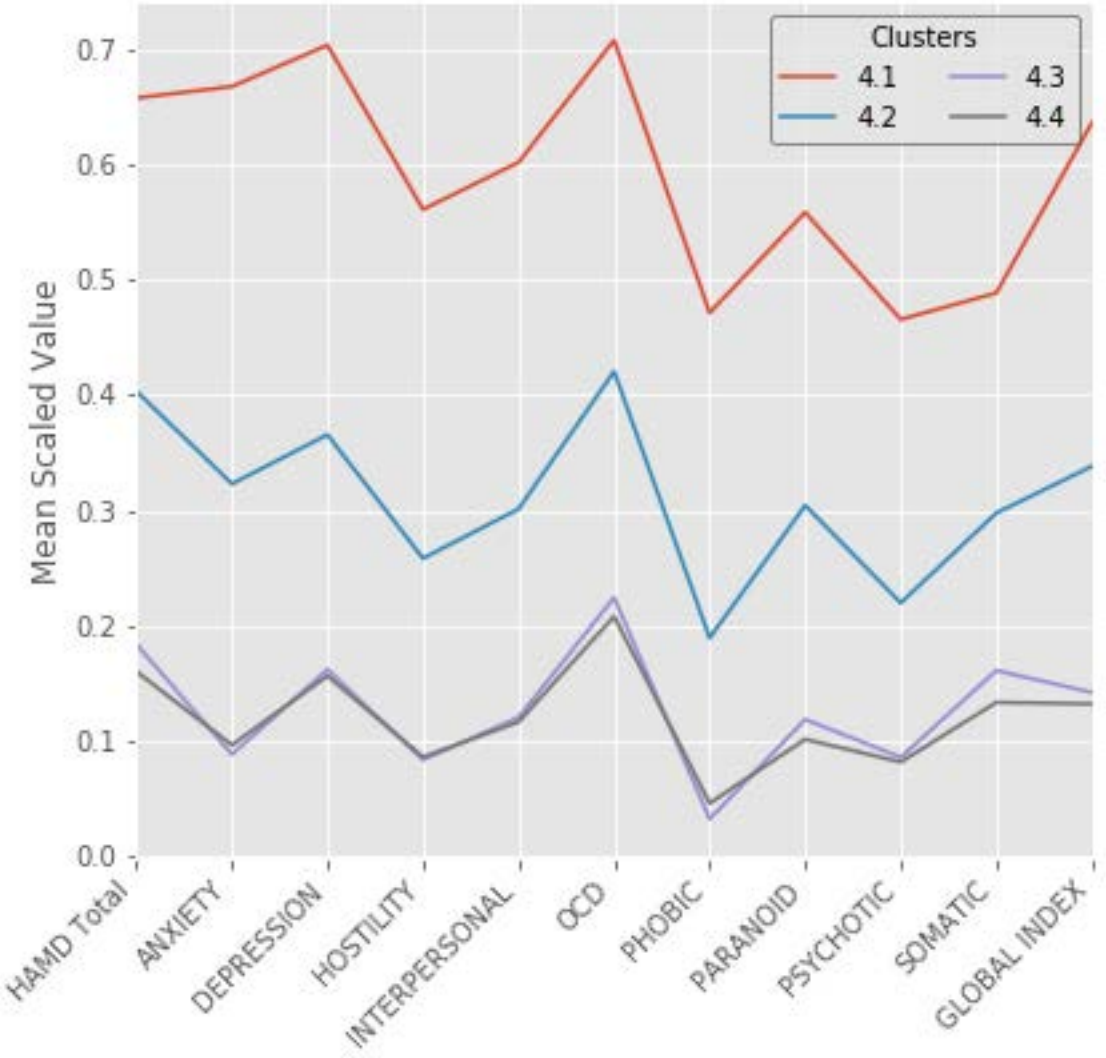

Figure 6.2b. Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 in the CS Dataset (k=4)

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	Cluster 3 Mean	S.D.	Cluster 4 Mean	S.D.	ANCOVA P-value	Sig	F	Post Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=3,297)												
Executive Functioning 1	8.52E+00	2.95E+00	9.51E+00	3.34E+00	1.09E+01	2.92E+00	1.08E+01	3.12E+00	1.08E-04	***	7.26	3>1**,3>2*,4>1**
Executive Functioning 2	1.09E+01	2.45E+00	1.06E+01	2.90E+00	1.21E+01	2.95E+00	1.17E+01	2.97E+00	1.79E-04	***	6.84	3>1**,3>2**,4>2*
Learning 1	5.98E+00	1.15E+00	5.96E+00	1.94E+00	6.95E+00	1.91E+00	6.15E+00	1.83E+00	4.42E-02	*	2.73	3>1*
Learning 2	1.09E+01	2.94E+00	1.14E+01	2.61E+00	1.24E+01	2.35E+00	1.25E+01	2.82E+00	9.88E-05	****	7.29	3>1***,3>2**,4>1**
Learning 3	4.18E+01	9.68E+00	4.76E+01	1.07E+01	5.14E+01	9.50E+00	5.06E+01	1.02E+01	1.10E-05	****	8.93	2>1*,3>1****,4>1***
Memory 1	9.28E+00	2.76E+00	9.82E+00	3.64E+00	1.14E+01	2.90E+00	1.19E+01	3.19E+00	2.41E-04	***	6.62	3>1**, 3>2**,4>1*,4>2*
Memory 2	1.09E+01	2.54E+00	1.06E+01	3.28E+00	1.21E+01	2.71E+00	1.20E+01	2.86E+00	2.04E-04	***	6.75	3>1*, 3>2**,4>1*,4>2*
Performance IQ 1	5.82E+01	1.16E+01	6.60E+01	1.90E+01	6.65E+01	1.62E+01	7.00E+01	1.91E+01	7.02E-04	***	5.82	3>1*,4>1***
Performance IQ 2	9.52E+01	1.16E+01	9.99E+01	1.17E+01	4.19E+01	1.19E+01	4.48E+01	1.17E+01	2.39E-04	***	6.65	3>1*,4>2**,4>2*
Verbal IQ 3	1.51E+01	3.85E+00	1.59E+01	4.09E+00	1.69E+01	4.26E+00	1.75E+01	4.14E+00	1.18E-02	*	3.72	4>1*
Imaging Volumes (D.F.=3,122)												
Right Frontal Cortex	8.15E-02	8.58E-03	8.25E-02	8.77E-03	8.50E-02	5.41E-03	8.41E-02	5.75E-03	1.29E-02	*	3.68	N.S.
Left Frontal Cortex	8.17E-02	7.55E-03	8.20E-02	8.49E-03	8.49E-02	5.55E-03	8.34E-02	5.88E-03	8.99E-03	**	3.96	N.S.
Right Temporal Cortex	4.96E-02	5.27E-03	4.98E-02	5.92E-03	5.17E-02	4.01E-03	5.18E-02	3.98E-03	6.75E-03	**	4.17	N.S.
Left Temporal Cortex	4.99E-02	5.42E-03	4.96E-02	5.55E-03	5.15E-02	4.07E-03	5.15E-02	3.48E-03	1.92E-02	*	3.66	N.S.
Right Frontal White Matter	9.95E-02	9.60E-03	9.42E-02	1.44E-02	9.04E-02	7.30E-03	8.78E-02	7.10E-03	2.75E-03	**	4.85	2>4**
Left Frontal White Matter	9.29E-02	1.15E-02	9.34E-02	1.48E-02	9.00E-02	7.80E-03	8.74E-02	7.10E-03	1.01E-02	*	3.87	2>4**
Right Parietal White Matter	4.78E-02	6.97E-03	4.70E-02	7.84E-03	4.54E-02	3.65E-03	4.48E-02	3.91E-03	4.36E-02	*	2.75	N.S.
Left Parietal White Matter	4.57E-02	4.00E-03	4.57E-02	7.50E-03	4.34E-02	3.40E-03	4.35E-02	3.90E-03	1.45E-02	*	3.59	2>3*
Right Occipital White Matter	1.78E-02	2.40E-03	1.75E-02	2.40E-03	1.69E-02	2.10E-03	1.64E-02	1.90E-03	1.30E-02	*	3.68	2>4*
Left Occipital White Matter	1.78E-02	3.20E-03	1.78E-02	2.50E-03	1.72E-02	2.20E-03	1.63E-02	2.00E-03	3.11E-03	**	4.75	2>4**
Right Temporal White Matter	9.69E-02	5.90E-03	9.72E-02	8.60E-03	9.53E-02	4.10E-03	9.40E-02	2.40E-03	8.05E-03	**	4.04	2>4**
Right Parietal CSF	2.09E-02	4.90E-03	2.00E-02	4.70E-03	2.12E-02	5.50E-03	2.40E-02	4.80E-03	7.34E-05	****	7.6	4>2***,4>3**
Left Parietal CSF	1.96E-02	5.50E-03	1.95E-02	4.70E-03	2.02E-02	4.90E-03	2.29E-02	4.60E-03	1.57E-04	***	7.02	4>1*,4>2***,4>3***
Right Occipital CSF	4.98E-03	1.20E-03	4.69E-03	1.50E-03	4.57E-03	1.30E-03	5.58E-03	1.40E-03	6.48E-05	****	7.69	4>1**, 4>2**,4>3***,
Left Occipital CSF	3.95E-03	1.20E-03	4.50E-03	1.90E-03	4.49E-03	1.20E-03	5.53E-03	1.50E-03	1.06E-07	****	12.72	4>1***,4>2***,4>3****
Right Temporal CSF	1.45E-02	3.10E-03	1.42E-02	2.90E-03	1.45E-02	2.70E-03	1.57E-02	2.90E-03	4.50E-03	**	4.48	4>2*
Left Temporal CSF	1.34E-02	3.07E-03	1.31E-02	2.76E-03	1.34E-02	3.02E-03	1.44E-02	3.02E-03	3.07E-02	*	3.02	N.S.
Cortisol Measures (D.F.=3,287)												
Baseline AM Cortisol Day 1 of DST	3.05E+00	7.10E-01	3.27E+00	6.50E-01	3.35E+00	5.00E-01	3.44E+00	4.60E-01	3.09E-02	*	3.02	4>1*
Percent Suppression Baseline Cortisol to DST(D1-D2/D1)	8.10E+01	2.22E+01	7.38E+01	2.99E+01	8.63E+01	1.92E+01	6.16E+01	3.25E+01	6.95E-04	***	5.91	3>1*,4>1***

D.F., degrees of freedom
S.D., standard deviation
TK, Tukey-Kramer
*, p<.05; **, p<.01; ***, p<.001; ****, p<.0001.
N.S., not Significant

Table 6.2. Biomarker Differences Across Clusters in the CS Dataset for k=4

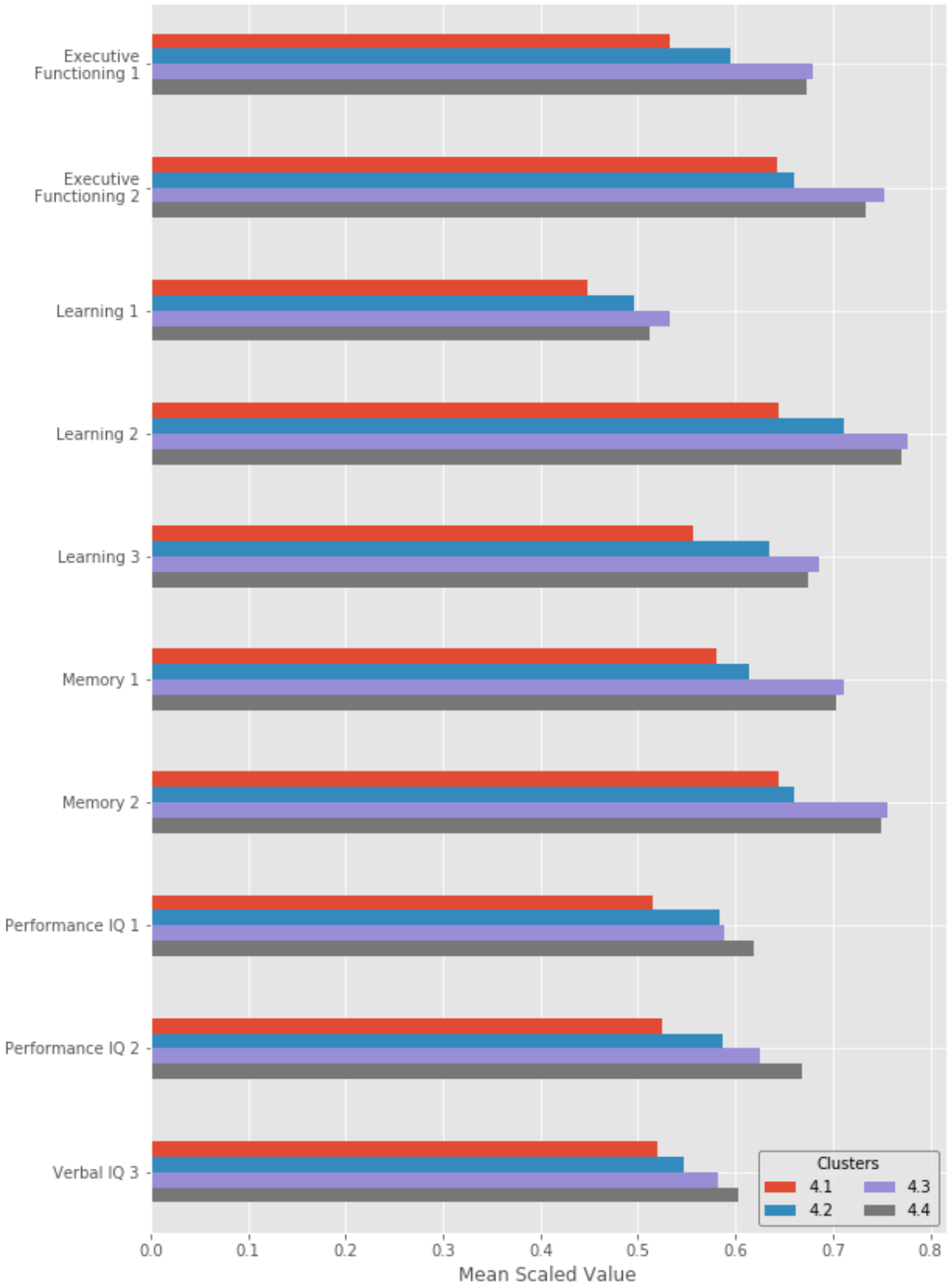

Figure 6.3a. Neuropsychiatric Markers with Significant Differences Across Clusters (k=4) in the CS Dataset

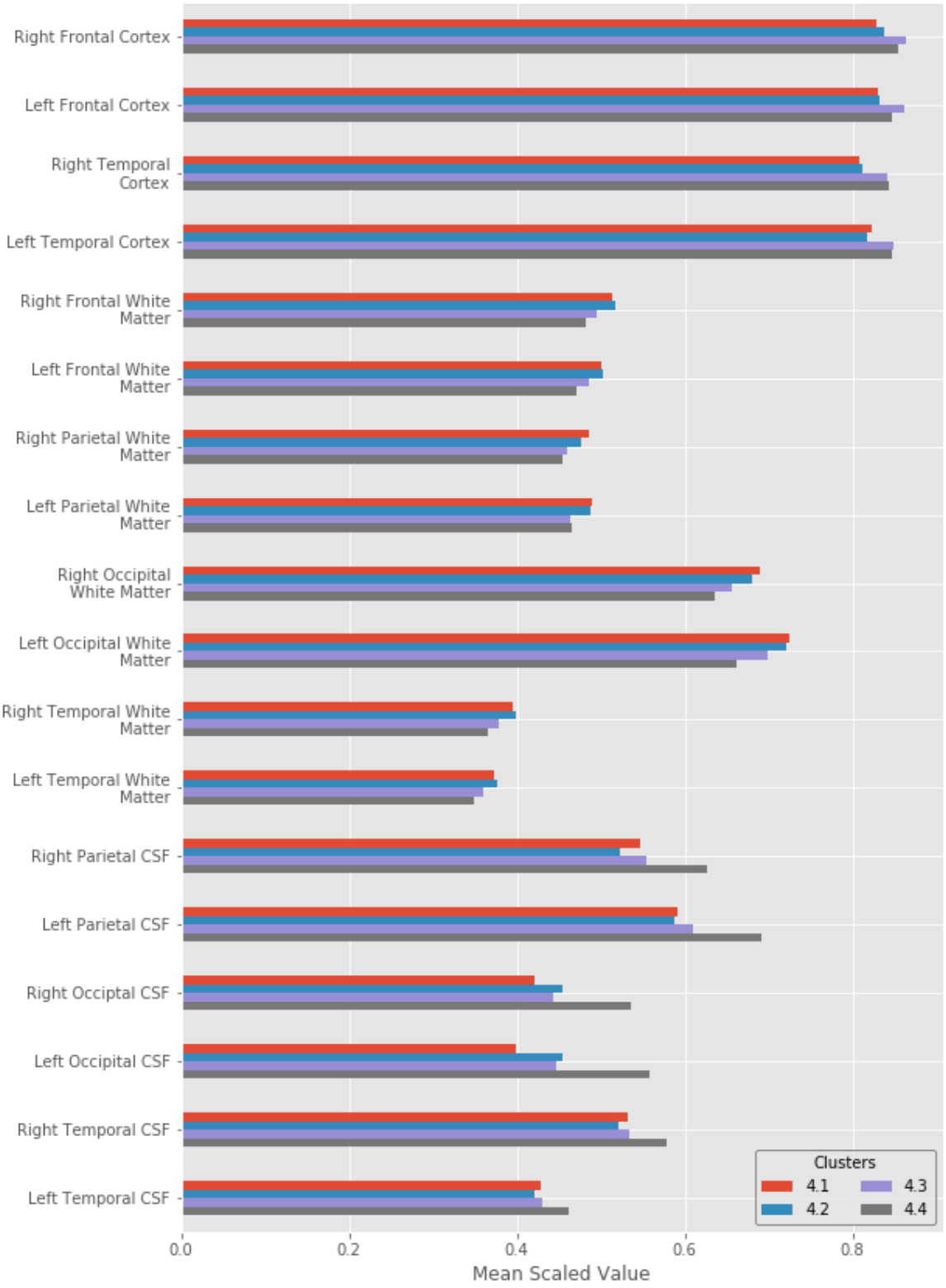

Figure 6.3b. Imaging Markers with Significant Differences Across Clusters (k=4) in the CS Dataset

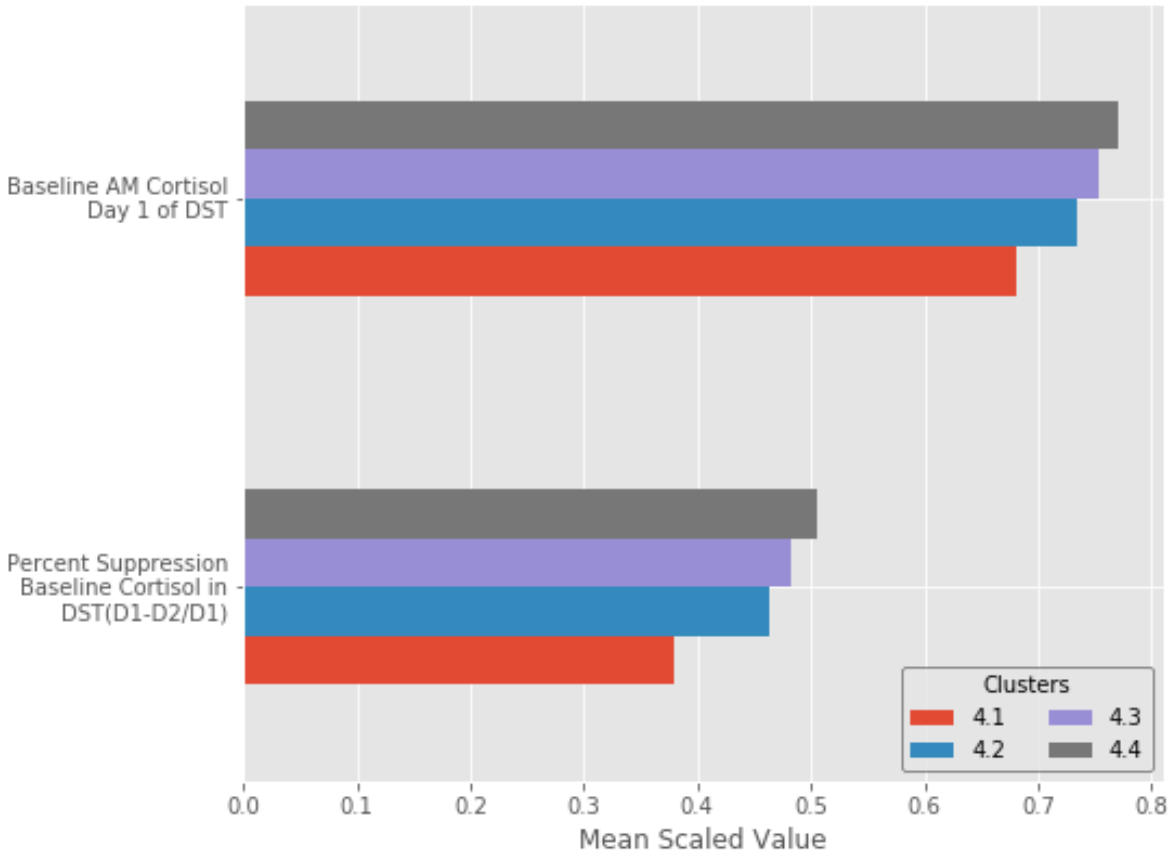

Figure 6.3c. Endocrine Markers with Significant Differences Across Clusters (k=4) in the CS Dataset

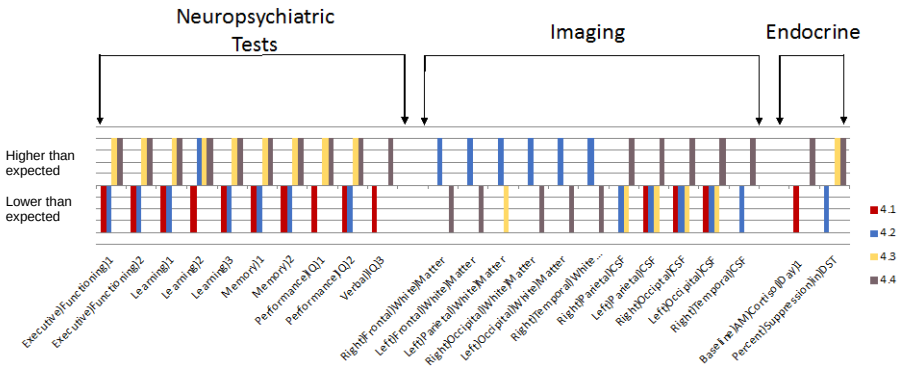

Figure 6.4. Rich Clinical Phenotype of CS k=4 Solution

To contrast the differences between using more granular psychiatric symptom data in the clustered CS with the DSM PTSD diagnosis, individuals with a diagnosis of PTSD in the Combined Dataset are compared with those without the diagnoses below. Figures 6.5a-b show the different symptom profiles of these two groups. Unlike the previous four-cluster solution where symptom profiles are similar or overlap in parts of the clinical inventories, in the PTSD vs. No PTSD inventories (Figures 6.2a and 6.2b), the two groups are clearly delineated, with the PTSD group having higher values across all the items in the clinical inventories (Figures 6.5a and 6.5b).

Table 6.3 and Figure 6.6a show that in all of the neuropsychiatric tests where there are significant differences between groups, the PTSD group had lower or "worse" scores on the neuropsychiatric tests. Figure 6.6b shows the same trends that are seen in all of the CS clustering solutions (Figures 6.3b, Appendix D4), with subjects with PTSD having lower mean cortical gray and CSF matter volumes, and higher mean white matter cortical volumes.

Perhaps the most interesting absence of findings in the PTSD vs. No PTSD analyses is the lack of differences in both the baseline Day 1 a.m. cortisol levels and the Day 2 cortisol levels following dexamethasone administration via the DST. This finding is present in the CS analyses for the k=2, k=3, and k=4 cluster solutions (Figure 6.3c, Table 6.2, Appendix D4), even though the groups in the k=4 cluster solutions were relatively small in size as compared to the PTSD vs. No PTSD analyses, and therefore had less power to detect a difference across groups. Essentially, the CS four-cluster solution was able to delineate two subgroups of individuals with PTSD diagnoses: one with significantly lower baseline cortisol levels, and one with post-DST cortisol hypo-suppression, as compared with two subgroups of individuals without PTSD. Further ramifications of this finding will be elaborated on in the discussion section of this paper.

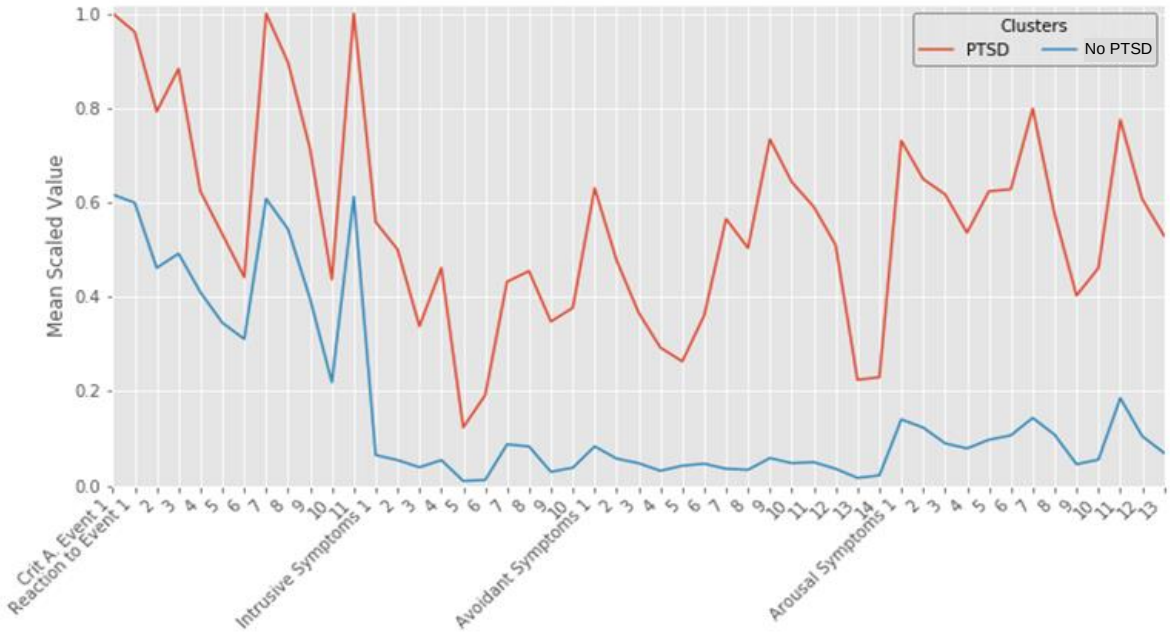

Figure 6.5a. Symptom Profiles for All Items in the Clinician Assessment for PTSD vs. No PTSD in the CS Dataset

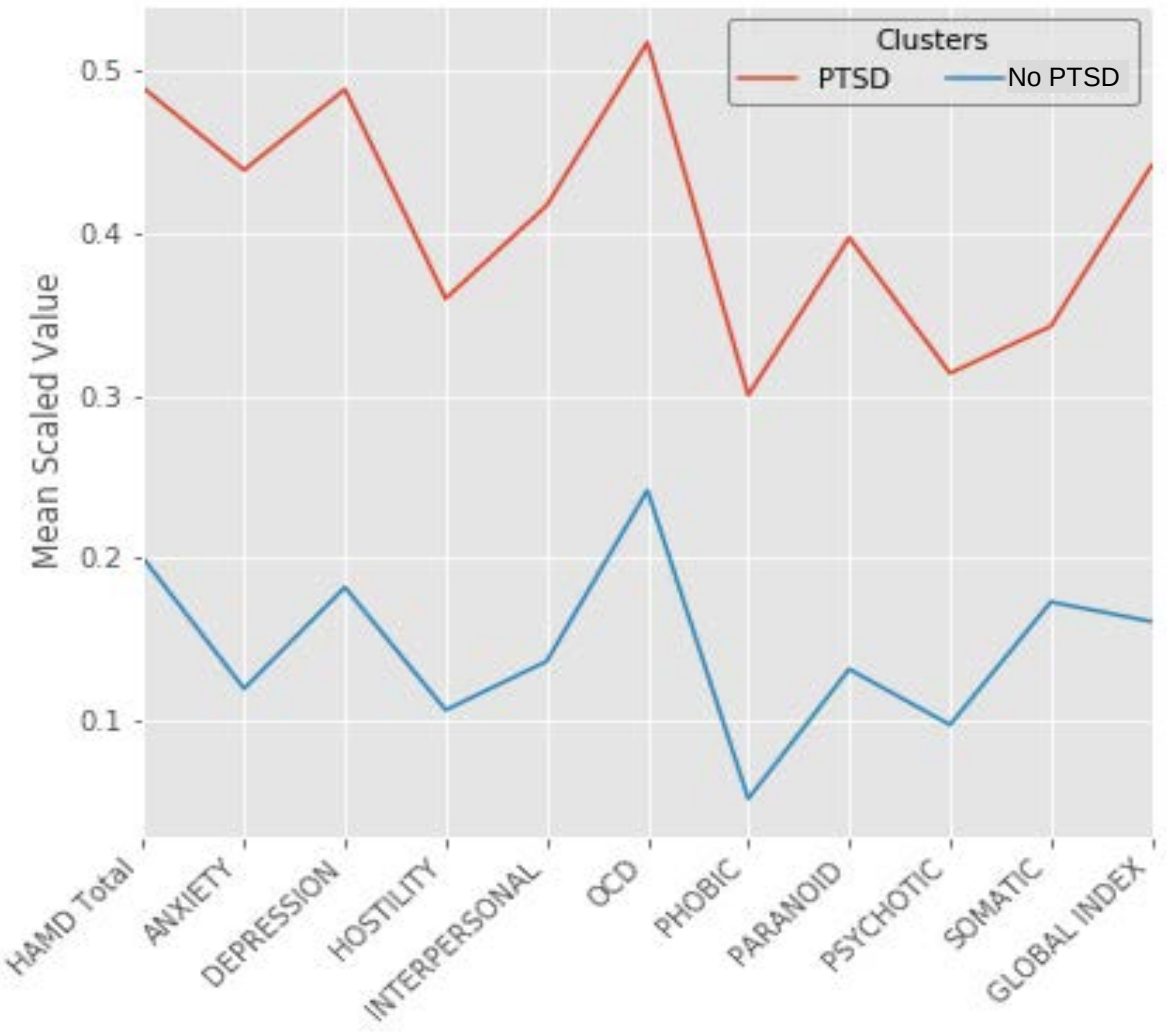

Figure 6.5b. Symptom Profiles for All Summary Items in the BDI and SCL-90 in the CS Dataset for PTSD vs. No PTSD

Table 6.3 Biomarker Differences Across Clusters in the CS Dataset Between PTSD vs. No PTSD

Test	PTSD (Clus.1) Mean	S.D.	No PTSD (Clus.2) Mean	S.D.	ANCOVA P-value	Sig	F	Post Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=1,299)								
Learning 1	5.50E+00	1.70E+00	6.34E+00	1.85E+00	4.54E-04	***	4.92	2>1*T
Learning 2	1.08E+01	2.71E+00	1.23E+01	2.56E+00	1.22E-05	****	4.82	2 > 1*****
Learning 3	4.41E+01	1.03E+01	5.12E+01	9.76E+00	1.31E-07	****	8.19	2 > 1*****
Executive Functioning 1	9.16E+00	2.92E+00	1.07E+01	3.10E+00	2.10E-04	***	7.84	2 > 1****
Executive Functioning 2	1.04E+01	2.73E+00	1.18E+01	2.69E+00	1.74E-04	***	9.75	2 > 1***
Memory 1	9.54E+00	3.19E+00	1.12E+01	3.15E+00	7.65E-05	****	10.09	2 > 1*****
Memory 2	1.03E+01	2.96E+00	1.20E+01	2.83E+00	1.44E-05	****	12.67	2 > 1*****
Performance IQ 1	6.23E+01	1.30E+01	6.80E+01	1.48E+01	1.28E-03	**	N.S.	2 > 1***
Performance IQ 2	3.74E+01	1.14E+01	4.28E+01	1.20E+01	2.19E-04	***	7.1	2 > 1***
Verbal IQ 3	1.54E+01	3.81E+00	1.71E+01	4.24E+00	2.87E-03	**	N.S.	2 > 1***
Verbal IQ 4	1.86E+01	4.88E+00	1.98E+01	4.27E+00	3.71E-02	*	4.37	2 > 1**
Imaging Volumes(D.F.=1,223)								
Right Frontal Cortex	8.17E-02	8.98E-03	8.46E-02	5.56E-03	2.15E-04	***	14.15	2 > 1***
Left Frontal Cortex	8.16E-02	8.49E-03	8.42E-02	5.72E-03	9.69E-04	***	11.18	2 > 1***
Right Temporal Cortex	4.89E-02	5.81E-03	5.19E-02	3.69E-03	2.28E-07	****	28.52	2 > 1*****
Left Temporal Cortex	4.91E-02	5.58E-03	5.16E-02	3.78E-03	1.04E-05	****	20.35	2 > 1*****
Right Frontal White Matter	9.51E-02	1.40E-02	8.93E-02	7.13E-03	5.75E-05	****	16.83	1 > 2****
Right Parietal White Matter	4.77E-02	7.99E-03	4.51E-02	3.75E-03	6.74E-04	***	11.89	1 > 2***
Left Parietal White Matter	4.61E-02	7.27E-03	4.34E-02	3.31E-03	1.67E-04	***	14.67	1 > 2***
Right Occipital White Matter	1.77E-02	2.40E-03	1.67E-02	2.05E-03	8.93E-04	***	11.34	1 > 2***
Left Occipital White Matter	1.80E-02	2.74E-03	1.68E-02	2.15E-03	8.20E-04	***	11.51	1 > 2***
Right Temporal White Matter	3.76E-02	8.4E-03	3.47E-02	3.37E-03	3.90E-04	***	12.97	1 > 2***
Left Temporal White Matter	3.67E-02	8.79E-03	3.46E-02	3.28E-03	6.89E-03	**	7.44	1 > 2***
Right Parietal CSF	2.02E-02	4.80E-03	2.22E-02	5.35E-03	7.37E-03	**	7.32	1 > 2**
Left Parietal CSF	1.94E-02	5.06E-03	2.13E-02	4.90E-03	8.64E-03	**	7.01	2 > 1**
Right Occipital CSF	4.44E-03	1.46E-03	5.00E-03	1.38E-03	1.05E-02	*	6.66	2 > 1**
Left Occipital CSF	4.21E-03	1.35E-03	4.91E-03	1.38E-03	8.65E-04	***	11.4	2 > 1*
Right Temporal CSF	1.42E-02	3.06E-03	1.50E-02	2.82E-03	2.20E-02	*	5.32	2 > 1***

D.F.: degrees of freedom
S.D.: standard deviation
N.S.: not significant
TK: Tukey-Kramer
*, p<.05; **, p<.01; ***, p<.001, **** p<.0001, ***** p<.00001
T, PTSD is Cluster 1; No PTSD is Cluster 2

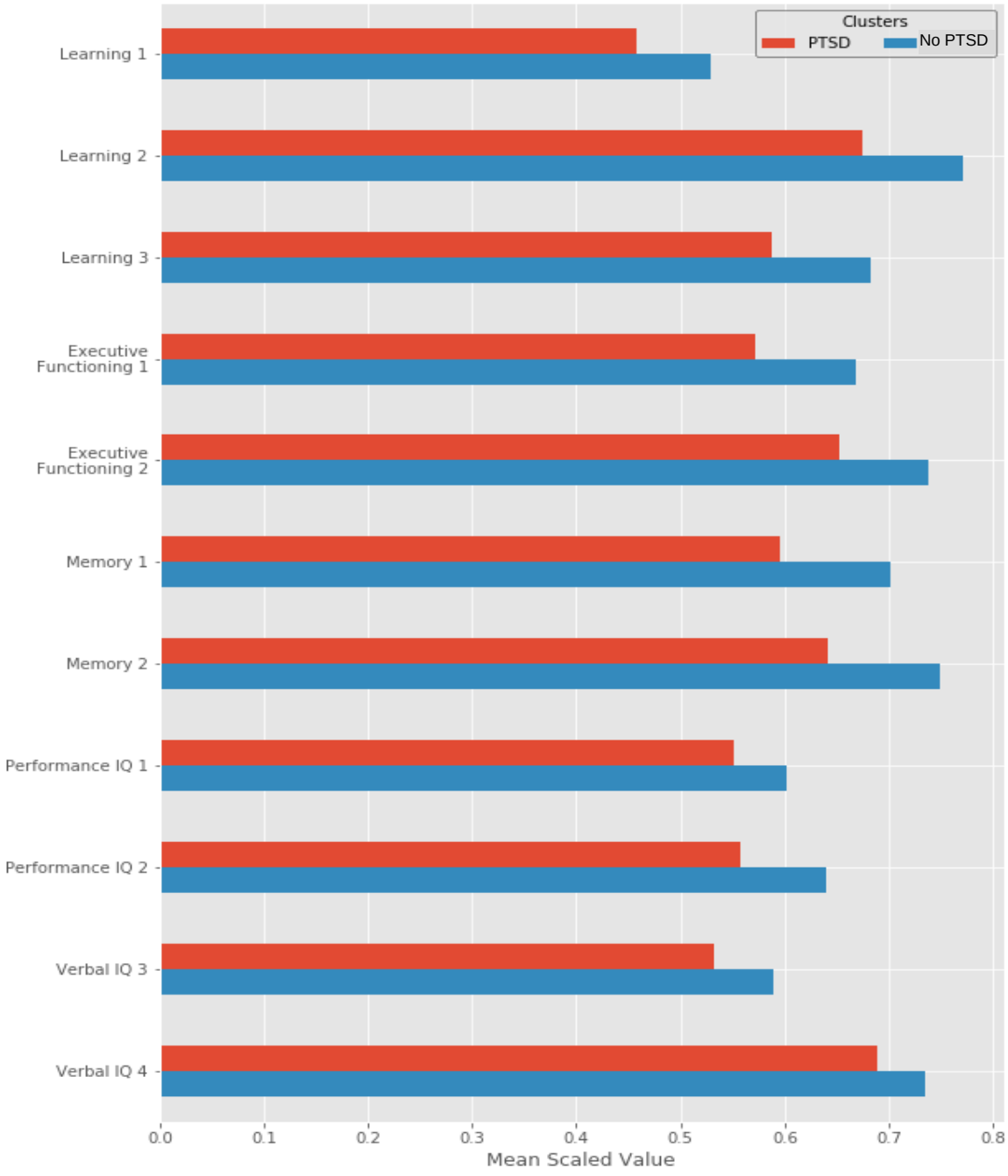

Figure 6.6a. Neuropsychiatric Markers with Significant Differences in CS Dataset Between PTSD vs. No PTSD

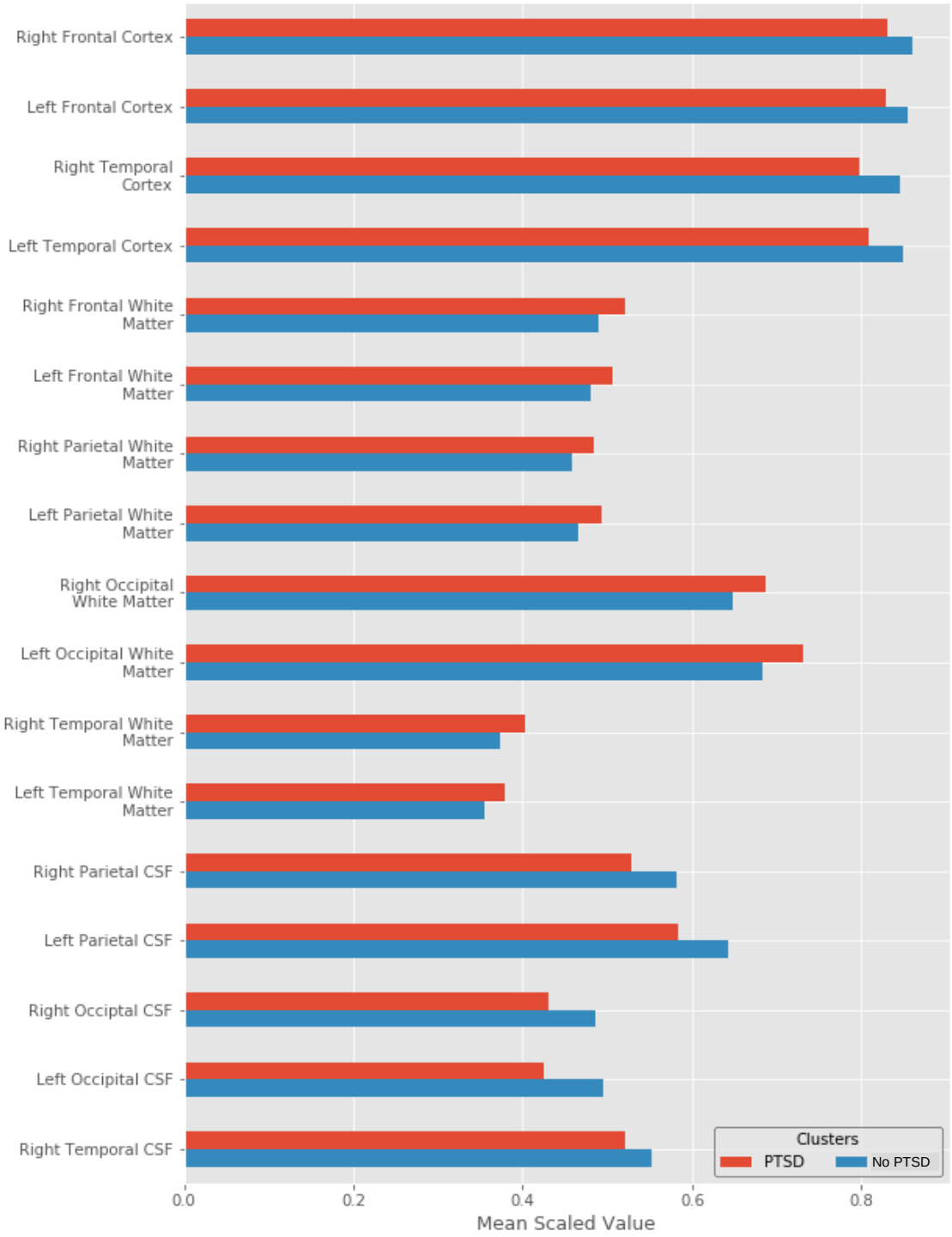

Figure 6.6b. Imaging Markers with Significant Differences in CS Dataset Between PTSD vs. No PTSD

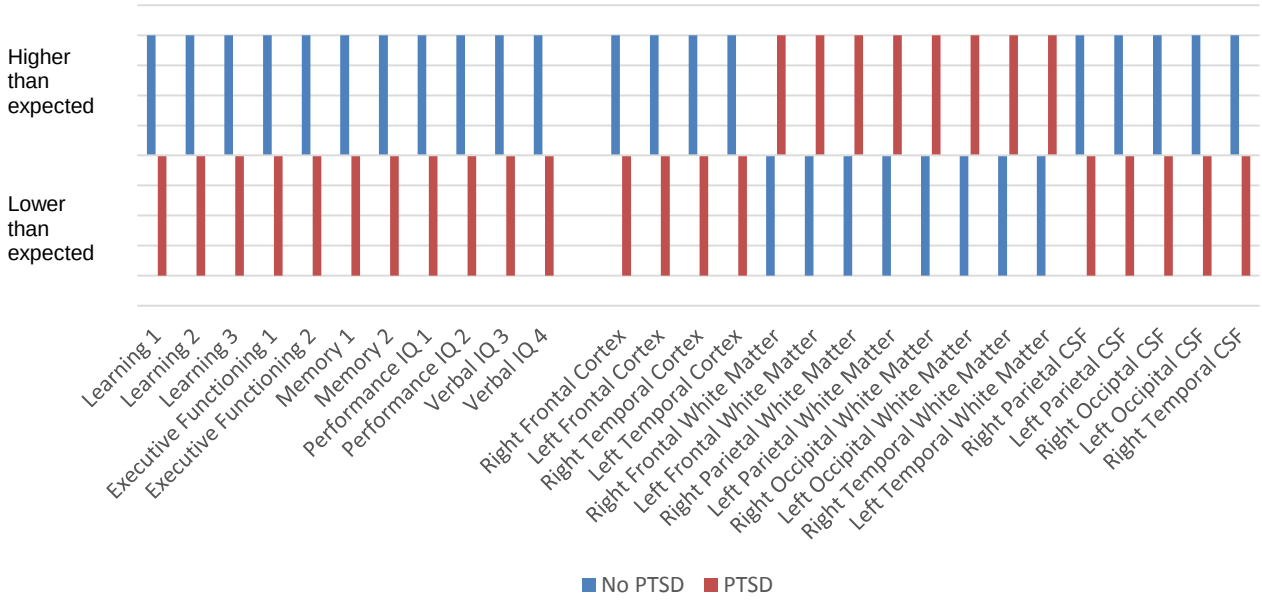

Figure 6.7. Rich Clinical Phenotype of the PTSD vs. No PTSD Analyses in the CS Dataset

Chapter 7: Discussion

The overarching question addressed in this paper is whether psychiatric symptom–level data, in contrast to DSM syndrome–level data, can be successfully used to inform a biologically based taxonomy of mental health disorders. As discussed in the Introduction, problems with the current classification system are widely documented, including the lack of specificity and binary nature of DSM syndromes, and the subsequent absence of identified biological correlates with clinical utility (Craddock & Owen, 2010; Jablensky, 1999).

As the DSM has been shown to have severe shortcomings when used to phenotype psychiatric populations to identify biological correlates of psychiatric illness, it has been suggested that it essentially be renounced, and a new taxonomy developed based on biological constructs (Insel, 2014). Yet there will be great challenges when transitioning from the widely used current classification system to using a fundamentally different one based on measurable biological constructs, especially in clinical practice (Cuthbert, 2014; McCoy et al., 2015; Sharp et al., 2016). The DSM is based only on self-reported symptoms or observed signs in patients with psychological distress, yet the current working proposed RDoCs matrix weighs heavily toward biological markers and measurable neuropsychiatric tasks, with limited use of self-report symptoms (Cuthbert & Insel, 2013). However, self-report symptoms are still the primary data used in clinical psychiatric practice to formulate diagnoses and to determine therapeutic interventions. Additionally, the importance of patients' self-presentations is supported by their continued place in clinical work in other medical domains to create differential diagnoses, and at least in part to inform the diagnostic workup, albeit often in the context of measureable signs (e.g., chest pain in the context of abnormal vital signs) (Goodman, Gilman, Brunton, Lazo, & Parker, 2006).

This leads to the question of how patient-reported symptoms will be used in clinical psychiatric practice in the age of precision medicine. This question is particularly important in the mental health domain for several practical and therapeutic reasons. Practicing psychiatrists and training programs use a current diagnostic system based completely on patient-reported symptoms and clinician-observable behaviors (American Psychiatric Association. & American Psychiatric Association. DSM-5 Task Force.). Remuneration from insurance companies and Center for Medicare and Medicaid Services also relies on the DSM, which is cross-referenced to the International Classification to Disease (ICD) diagnoses (Meyer, 2011). Perhaps most importantly, the therapeutic relationship between the psychiatrist and the patient relies on the verbal interaction where the patient relays their distressing symptoms to clinicians and, as a team, they use multiple therapeutic modalities, including medication management and psychotherapy, to work together toward decreasing the patient's distressing symptoms (Gabbard, 1990; Goodman et al., 2006). While the mechanisms of healing through psychotherapy are far from well understood, the efficacy of psychotherapeutic treatments are well documented, and an important aspect of the therapeutic relationship is believed to be the interactions that occur as the clinician learns about their patient's subjective experience (Gabbard, 1990; Hales et al.; Shedler, 2010). It is difficult to conceive of how this transaction will transpire if patient evaluations are predominantly conducted through cognitive and behavioral tasks that lack genuine human interaction. Furthermore, in line with human experience, individuals often are able to better define and understand their experience through interaction with others, including their treatment providers. It is thought that this increase in understanding of the patient's psychological narrative, and the insight that accompanies it, is vital to improvements in the mental state of afflicted individuals (Gabbard, 1990).

For these reasons, it is imperative to envision how symptoms will be used to inform a new psychiatric taxonomy that is being developed from a scientific perspective, grounded in objectively observable and quantifiable measures. National initiatives are under way that are focused on ascertaining clinical and biological data in large samples of individuals with the goal of identifying clinical and biological markers, which will ultimately facilitate this translation (Collins & Varmus, 2015; Gaziano et al., 2016). In line with these efforts, the approach in this study has been to attempt to use the individual self-report symptom data from three different psychiatric inventories, and determine if machine-learning algorithms can discover groups of patients with similar symptom profiles that differ with respect to biological markers.

Although the size of the datasets used in this study are suboptimal for this type of analysis, these datasets are multimodal, including clinical inventories, neuropsychiatric data, imaging data, endocrine data, and genetic data. At this time, there remains a paucity of large multimodal datasets, especially those including imaging data, available in psychiatry thus far. Maximizing the use of datasets that have already been ascertained using significant resources to identify subgroups with increased heterogeneity will ideally inform larger scale symptomatic data analysis and/or acquisition. There are numerous attempts to ascertain large multi-modal datasets (Collins & Varmus, 2015; Gaziano et al., 2016; Turner, 2014), which will hopefully be facilitated by the far reach of the internet, that have already begun to allow for large-scale remote data acquisition, the widespread adoption of EHRs, and natural language-processing methods that may be able to glean symptom data from narrative text (McCoy et al., 2015; Perlis et al., 2012).

The hypothesis underlying the approach of using all available Likert scale self-report symptom data in these datasets to identify groups of patients with clinically homogenous

presentations is that these analyses might yield different results than prior studies using DSM syndromes as phenotypes, due to the increased granularity and breadth of the data used, which might increase power to identify groups with greater clinical homogeneity with biological correlates. Cluster analysis was employed to define data-driven psychiatric phenotypes in the form of symptom profiles that might aid in explaining the lack of consistency in the psychiatric literature concerning biomarkers associated with DSM diagnoses.

Given the overall structure of the DSM, it was expected that if patients were clustered using a wide array of symptoms, groups would emerge that had qualitatively different profiles, e.g., that some groups would have a higher level of certain types of symptoms, and others would have higher levels in other items or types of symptoms. This expectation is consistent with both theories of psychopathology as well as the current DSM classification system used in clinical psychiatry today (American Psychiatric Association. & American Psychiatric Association. DSM-5 Task Force.; Gabbard, 1990). The findings with these data, however, was that individuals were clustered by intensity level across all symptoms, with extremely little overlap from one profile to the next. Given the review of the literature in Chapter 2 of this study, perhaps this finding should have been less surprising.

While initially the results were discouraging given the expectations of symptom profiles that differed qualitatively, and hence had more visual overlap, the clinical and biological marker patterns associated with these clusters do in fact provide insight that can increase the understanding of psychiatric distress. Individuals who report higher levels of symptom distress in this study differed from those with low reports of distress across a variety of clinical, neuropsychiatric, imaging, and endocrine measures that were not used to create the symptom profile clusters, providing external validation that the clusters defined by self-report symptoms

alone are able to identify groups with increased clinical and biological homogeneity. This was despite the very small sample size, as well as the considerable differences in the design of the two studies that were merged to provide the largest, and hence most highly powered, sample that was analyzed.

The first external clinical marker analyzed was alcohol abuse and dependence, and indeed there is a greater incidence of alcohol abuse in clusters with higher symptom profiles than in those with lower symptom profiles (Tables 5.1, 6.1, D2.1, D3.1). While not surprising that individuals who experience more distressing symptoms also use more alcohol, no data concerning alcohol or substance use was included in the dataset employed in the cluster analyses. These results are consistent with larger population studies that suggest that individuals use alcohol and other substances to self-medicate for distressing symptoms, and that mental illness is more prevalent in individuals who abuse alcohol than in the general population (Regier et al., 1990; Tsuang, Tohen, & Jones, 2011). Additionally, these results support prior authors' assessments that evaluating and classifying individuals separately for alcohol-use disorders and other DSM diagnoses may be counterproductive, as this approach has led to many studies that investigate cohorts with alcohol use while excluding those that meet criteria for another DSM disorder (Pettinati, O'Brien, & Dundon, 2013). This not only ignores subclinical mental health diagnoses in individuals being studied, and hence possibly important contributing confounders, but also has led to a paucity of evidence on how to treat comorbid alcohol abuse and mental distress, despite the high utilization of services found in patients identified as having dual diagnoses (Connor, Haber, & Hall, 2016; Hall, Degenhardt, & Teesson, 2009). Studies of alcohol use that don't take into account the self-report symptoms of psychiatric patients are limiting progress in this important domain of public health (Iovieno, Tedeschini, Bentley, Evins, &

Papakostas, 2011; Lev-Ran, Balchand, Lefebvre, Araki, & Le Foll, 2012; Murthy & Chand, 2012).

Clusters of individuals with higher symptom profiles also had on average higher incidences of history of childhood abuse, providing further external validation that self-report symptom data may be able to identify individuals with different psychosocial histories and experiences. These results can be viewed as consistent with prior reports in the literature that adverse childhood events are associated with mental health issues in adulthood (Schilling, Aseltine, & Gore, 2007). Multiple studies in the literature have found associations between childhood trauma and a wide range of psychiatric disorders, including mood disorders (Chapman et al., 2004; Heim & Nemeroff, 2001), anxiety disorders (Heim & Nemeroff, 2001; Lochner et al., 2002), trauma- and stressor-related disorders (Brewin, Andrews, & Valentine, 2000; Ozer, Best, Lipsey, & Weiss, 2003), personality disorders (Afifi et al., 2011), psychotic disorders (Alvarez et al., 2011; Bendall, Jackson, Hulbert, & McGorry, 2008), and alcohol and substance use disorders (Dube, Anda, Felitti, Edwards, & Croft, 2002; Dube et al., 2003). Furthermore, findings that an increased proportion of individuals in the higher symptom clusters have both comorbid alcohol use and a history of childhood trauma may point to a potential use for symptom profiles within the public health domain, as individuals with severe mental illness and a trauma history have increased rates of utilization of mental health services (Schneeberger et al., 2017).

With respect to neuropsychiatric testing, the overall trend was that individuals with higher symptom profiles scored worse on neuropsychiatric tests, although the details are not completely straightforward. There was no single neuropsychiatric measure represented in clusters with the most severe symptom profile across analyses. Both Executive Function and

Performance IQ were significantly worse in clusters with higher symptom profiles in both the GW and GS analyses, and this significance increased with the larger CS dataset. Learning was impaired in MS clusters with higher symptom profiles, and the significance of these inverse relationships increased in the CS dataset as well. Impairments in memory in contrast to cluster symptom profile were significant to the same extent in the GW study as with the CS dataset. Consistent with these findings, numerous publications have demonstrated a decrease in function in these three neuropsychiatric domains in a wide range of psychiatric populations, including individuals with PTSD, depression, and alcohol use disorders (Austin, Mitchell, & Goodwin, 2001; Goodkind et al., 2015; McTeague, Goodkind, & Etkin, 2016; Samuelson, 2011; Samuelson et al., 2006; Snyder, Miyake, & Hankin, 2015; Trivedi, 2006).

It is possible that the lack of replication of findings in the two individual datasets concerning symptom profiles and neuropsychiatric domains can in part be explained by the difference in populations in the two studies, where the Gulf War study had a minority of individuals with PTSD (33/292) and other DSM diagnoses, while the MIRECC had a 2x2 design with equal groups of PTSD+/ETOH+, PTSD+/ETOH-, PTSD-/ETOH+, PTSD/ETOH- (e.g., 57/130 with PTSD). Furthermore, the small size of both of the studies also suggests that the lack of replication of findings from the analysis of the GW dataset to the MS dataset, and vice versa, may have been due to a lack of power.

It remains extremely interesting that the neuropsychiatric domains became more significant in the CS dataset, obviously the group with the most power, than in any of the other analyses. In fact, depending on how stringently a multiple testing correction factor is applied, some of these measures remain statistically significant with a Bonferroni correction for all the 50 tests done per cluster solution (Executive Functioning and Memory), and some which remain

significant even if corrected for all the tests done across all cluster solutions (Learning), which is likely too stringent given that the datasets and cluster solutions overlap, so each test arguably should not be viewed as independent when applying a multiple testing correction (approximately 50 biomarkers across four cluster solutions, and four datasets : N=800) (Garamszegi, 2006). These results provide additional support for our hypotheses that using more highly granular symptom-level data may help to identify clinical groups that are more robustly associated with differences across biomarkers, as compared with using DSM syndromes as phenotypes.

With respect to structural MRI differences, again multiple findings have been published showing differences in individuals with PTSD in their brain structure (Karl et al., 2006) as well as with individuals with depression (Soares & Mann, 1997) and alcohol dependence (Welch, Carson, & Lawrie, 2013) that have not been consistently replicated in the literature. In this study, we found several overall patterns that emerged between increased symptom profiles and changes in structural MRI volumes, most notably, frontal and temporal cortical volumes (grey matter) were smaller in clusters with higher symptom profiles. In the CS dataset, bilateral frontal grey matter and bilateral temporal grey matter were reduced in clusters with higher symptom profiles. Furthermore, the CS k=5 cluster solutions showed that this inverse association between higher cluster symptom profiles and reduced bilateral temporal grey matter was very significant ($p<.0001$), and would still be significant even after accounting for all multiple tests conducted. The most robust clusters in the GS and MS datasets did not have this inverse association, although the MS k=2 did. The GS had several cluster solutions with inverse relations between symptom profile intensity and smaller bilateral occipital grey matter, and the GW k=3 and k=4 cluster solutions also showed a significant relationship between reduced parietal grey matter and

increased symptom profiles, but these relationships were not present in the MS dataset, or the CS.

Numerous other studies in a multitude of psychiatric populations have shown reductions in cortical volumes as well. Smaller cerebral cortical grey matter volumes have been found in subjects with PTSD as compared with controls (Woodward, Schaer, Kaloupek, Cediel, & Eliez, 2009). A recent study using a path model showed that PTSD predicted Metabolic Syndrome in 274 combat military veterans, which was associated with reduced cortical thickness in several cerebral regions, including the bilateral temporal lobes and the left occipital cortex (Wolf et al., 2016), adding to findings in other studies showing associations between PTSD and reduced cortical thickness in frontal and temporal lobes (Bing et al., 2013; Geuze et al., 2008; Li et al., 2014). Evidence for smaller frontal and temporal lobes were also identified in a meta-analysis of imaging studies in mood disorders (Drevets, 2000; Soares & Mann, 1997). There are also results in the literature that support our findings that clusters with higher symptom profile scores along with neuropsychiatric memory impairment are associated with reduced cortical thickness. One study showed that reduced cortical thickness was associated with worse performance on memory measures in individuals with PTSD (Geuze et al., 2008). Another showed that veterans with PTSD and worsening clinical symptoms over time had accelerated atrophy in the frontal and temporal lobes, which was associated with a decline in verbal memory (Cardenas et al., 2011).

In contrast to the frontal and temporal cortical volumes, which were smaller in clusters with higher symptom profiles, white matter volumes in these clusters tended to be larger. In the largest CS dataset, for all of the cluster solutions, white matter volumes were larger bilaterally in all four lobes in the brain (Table 6.2, Appendix D4), and in the most robust cluster (k=4; e.g., the highest ARI and most stable bootstrapping results), all but the left temporal white volume matter

was larger ($p<.05$) in clusters with higher symptom profiles (Table 6.2). There were trends with large white matter volumes in the left occipital white matter in the GW and GS, and right occipital white matter in MS, but none were significant at $p<.05$. However, right lenticular nuclei white matter and right caudate nuclear white matter volumes were significantly larger in clusters with higher symptom profiles in several of the GW and GS solutions, including those that were assessed as most robust at k=5 in the GW. Unfortunately, imaging data for the lenticular and caudate nuclear structures are not available for the MS and the CS dataset, so we were unable to attempt to replicate or expand on these results. These findings are interesting when viewed in the context of other studies that have shown white matter alterations in individuals with psychiatric disorders. A 2013 meta-analysis of white matter findings in PTSD showed both decreased and increased white matter volumes in patients with PTSD as compared with controls (Daniels, Lamke, Gaebler, Walter, & Scheel, 2013). The decreased volumes were found in the corpus callosum, an area not evaluated in our dataset. One study in abused children vs. non-abused children did show an increase in white matter volumes in the cerebellum and prefrontal cortex (Hanson et al., 2010). A meta-analysis of structural neuroimaging in mood disorders identified the best replicated finding was an increased rate of white matter and periventricular hyperintensities (Soares & Mann, 1997). Several recent studies have also found differences in white matter measures including white matter hyperintensities (WMH) and fractional anisotropy (FA) in patients with PTSD. These measures are not consistently or clearly correlated white matter volume changes as measured in our dataset, however, so comparing findings in studies using different imaging methods is not within the scope of these analyses (Wardlaw, Valdes Hernandez, & Munoz-Maniega, 2015).

Additionally, sulcal CSF volumes were smaller in bilateral parietal, occipital, and temporal lobes in clusters with higher symptom profiles across most of the GW and GS clusters, including the most robust GW k=5 solution. These inverse correlations continued with more significance in the CS dataset, this time with values in the parietal and occipital lobes being very significant ($p<.00001$), and meeting criteria for significance even with a stringent correction for multiple testing. These findings are challenging to interpret, as relatively little research appears to have been done with sulcal CSF volumes and psychiatric disorders, although reduced sulcal CSF volumes were found in one study of patients with PTSD as compared with controls, supporting these findings (Woodward et al., 2007).

One final particularly interesting finding is that of a smaller right hippocampal volume, where in the GW k=5 cluster solution, one cluster with a higher symptom profile had a significantly smaller hippocampal volume than another cluster with a lower overall symptom profile. As can be seen from Table 5.2 and Figure 5.3b Cluster 5.2 has a significantly smaller right hippocampal volume than Cluster 5.3 ($p<.05$). Cluster 5.1 also has a mean right hippocampal volume that is smaller than that of Clusters 5.3, 5.4, and 5.5. Perhaps in part due to the small number of individuals in the clusters, this difference was not significant in post-hoc analyses. In our initial paper published on a subset of these data, the mean right hippocampal value for the cluster with the highest symptom profile was in fact nominally lower than that of the next most symptomatically affected cluster of individuals, but none of the volumes in this initial study reached a level of statistical significance (Ross et al., 2015). The difference between right hippocampal volume in two clusters is notable given that hippocampal volume has been a biomarker that has been widely focused on in studies looking for imaging correlates associated with PTSD, as well as other mood disorders. As can be seen, cluster 5.3 has a significantly

greater hippocampal volume than 5.2. When the BDI and SCL-90 summary scores were used in the GS as compared to the GW dataset, however, this difference decreased to a value with $p>.05$ ($p=.0637$). Due to the small sample size, it is not possible to determine whether having more granular symptom data actually increased the ability to find a difference in hippocampal size, or whether lack of power caused the lack of repetition of this finding in the GS dataset. However, it is of interest that the p-value for the difference in right hippocampal volume in the GW in Clusters 5.2 and 5.3 is at the same significance level ($p<.05$) using only age as a covariate, than the difference in hippocampal volume found in the same sample using linear modeling when individuals in the Gulf War study without a Criteria A event were excluded, and individuals with PTSD were compared to individuals without PTSD (Apfel et al., 2011). In comparing these two studies conducted within the same sample, unsupervised clustering using only self-report symptom data is able to identify cohorts of patients with smaller hippocampal volume that with and without PTSD and comorbid psychiatric disorders, without statistically adjusting for these disorders as in the Apfel paper. One possible interpretation of these findings is that the individuals who did not meet criteria for PTSD also had disorders that have been associated with smaller hippocampal size in the literature (Agartz, Momenan, Rawlings, Kerich, & Hommer, 1999; Bremner et al., 2000; Lee et al., 2016), and by including self-report symptom data not regularly ascertained and utilized in studies regarding DSM phenotypes, this analysis was able to group together individuals with similar levels of psychiatric distress across a wide range of symptoms, and thereby enrich our knowledge of a trans-diagnostic clinical phenotype that is associated with a smaller hippocampal volume. These findings using symptom profiles may also help to explain, in addition to a lack of power in most studies, why the association between PTSD, MDD, and alcohol abuse with a smaller hippocampal size is not always replicated (Karl

et al., 2006). The DSM binary phenotypes utilized in studies may result in both affected and comparison groups that are not homogeneous enough with respect to their overall level of distress, as demonstrated by their response across a wider range of self-report symptoms, and possibly this level of distress drives the correlation with smaller hippocampal size.

Finally, with regard to the neuroendocrine data, the most robust findings were that the baseline resting cortisol level is lower in individuals with higher symptom profiles in both the MS and CS samples. This finding was most notable in the CS dataset in the most robust k=4 cluster solution, where Cluster 4.1 had lower basal cortisol levels than Cluster 4.4 ($p<.05$). Several studies of individuals with PTSD have shown an association between lower waking cortisol level and PTSD (Neylan et al., 2005; Yehuda, Boisoneau, Mason, & Giller, 1993). In contrast, other studies have found increased basal cortisol levels in individuals with PTSD vs. controls (Savic, Knezevic, Damjanovic, Spiric, & Matic, 2012), and additional reports suggest that PTSD with comorbid major depression results in lower resting cortisol (Oquendo et al., 2003). Additionally, in the CS sample, Cluster 4.1 had significant morning hyposuppression of cortisol following the evening administration of dexamethasone, when compared with Clusters 4.3 ($p<.05$) and 4.4 ($p<.001$). This result was also found in the MS dataset for all cluster solutions. In the literature, there are inconsistent findings of both hypo- and hypersuppression of cortisol following the DST in patients with PTSD (de Kloet et al., 2006). Again, these findings suggest that using more highly granular self-report symptom data may be facilitating the identification of subpopulations of individuals who meet criteria for PTSD, along with other individuals in the sample who do not, who have both reduced basal cortisol and suppression following the DST. Therefore, these biological correlates may not be consistently identified in

studies of PTSD vs. controls, as they in exist in a population that overlaps with the PTSD group but is not equivalent.

Given the algorithm for classification of DSM disorders, it is logical that individuals with more symptoms are going to be classified as having more DSM disorders (comorbid disorders), as was found in our sample. It is interesting that individuals had higher symptom profiles not only across symptoms routinely measured to determine DSM diagnoses, but also for other symptoms that are not usually ascertained or reported in studies of the diagnoses predominantly used in these studies (e.g., PTSD, MDD, and alcohol abuse). Many symptom groups in the SCL-90, for example, interpersonal sensitivity and somaticism, are not routinely thoroughly evaluated in clinical work or research studies when seeking patients with disorders for which these symptoms are not seminal, including PTSD, MDD, and alcohol abuse.

It is conceivable that using a wider variety of symptom types in these cluster analyses enabled the identification of more robust clusters than when using only the PTSD symptoms. As mentioned in the chapter on workflow, the most robust clusters as determined by the ARI internal metric for cluster validity were those that were derived using data from the greatest breadth of symptoms in the GW dataset, without feature reduction. Unfortunately, only the sub-scores and not the complete symptom level data from the SCL-90 and Ham-D were available in the MIRECC study, so we were unable to test this hypothesis in the MS and the CS samples.

In an effort to show these findings in a concise graphical manner, a novel representation of a rich clinical phenotype has been presented that has the ability to compare different groups of patients across clinical and biological markers. With the GW dataset, the self-report symptoms are identified that are significantly higher or lower than their expected values based on the overall pattern of the symptom profiles. In both the GW and the CS, biological markers

	UNITS OF ANALYSIS							
DOMAINS/CONSTRUCTS	**Genes**	**Molecules**	**Cells**	**Circuits**	**Physiology**	**Behavior**	**Self-Reports**	**Paradigms**
Negative Valence Systems								
Acute threat ("fear")								
Potential threat ("anxiety")								
Sustained threat								
Loss								
Frustrative nonreward								
Positive Valence Systems								
Approach motivation								
Initial responsiveness to reward								
Sustained responsiveness to reward								
Reward learning								
Habit								
Cognitive Systems								
Attention								
Perception								
Working memory								
Declarative memory								
Language behavior								
Cognitive (effortful) control								
Systems for Social Processes								
Affiliation/attachment								
Social communication								
Perception/understanding of self								
Perception/understanding of others								
Arousal/Modulatory Systems								
Arousal								
Biological rhythms								
Sleep-wake								

Figure 7.1. Research Diagnostic Criteria (RDoC) Matrix (Cuthbert & Insel, 2013)

are identified that are significantly higher or lower than the means of those values in the other groups. While this mixture may be unusual, this representation is meant to facilitate a discussion in the literature as to how these findings may be integrated into the existing RDoC matrix, and for this purpose it is adequate.

The rich clinical phenotypes presented in this paper allow for differences across multimodal markers in groups of patients to be compared in a manner consistent with the RDoCs matrix. The RDoCs matrix was initially conceptualized as an actual matrix where five overall domains were identified in the first column: Negative Valence Systems, Positive Valence Systems, Cognitive Symptoms, Social Processes, and Arousal and Regulatory Systems, as shown in Figure 7.1. Within each domain, “constructs” are defined. For example, within Negative

Valence Systems, two of the constructs are Acute Threat ("fear") and Potential Threat ("anxiety"). The subsequent columns identify units of analysis for which support for these constructs have been, or potentially may be, identified. These broad groups of units represent increasing levels of biological complexity, from genes, molecules, and cells, to circuits and physiology, and finally behavior and self-reports. The self-reports identified in the existing RDoCs constructs include mostly Clinical Inventories (such as the Child Trauma Questionnaire used in this study), but also instances of individual symptoms, including amotivation, anhedonia, sleepiness, alertness, well-being, craving, mood, and fatigue. Additionally, it should be noted that neither this original RDoCs matrix nor the two subsequent iterations are meant to be exhaustive, and new constructs are expected to be added as the science progresses (Insel, 2014). Furthermore, not all constructs currently defined have elements across all units of analyses, although theoretically these will be identified in future studies.

The results in this study suggest that large amounts of individual symptom-level data may be used to identify groups with clinical and biological homogeneity, which may ultimately be conceptualized and investigated as constructs. While the Gulf War sample in this study with highly granular symptom data is unfortunately too small to ultimately delineate significant associations from spurious ones, this investigation of symptoms in a transdiagnostic population identified groups of individuals with more homogeneity biologically than when grouped by their DSM diagnoses. It can therefore by hypothesized that, for instance, individuals with specific symptom profiles have differences in symptoms, behaviors (neuropsychiatric tests are defined as behaviors in RDoCs), and circuits (e.g., a hippocampal circuit based on the absolute differences in hippocampal sizes). While the construct for the putative group identified in Cluster 5.2 in this

paper remains to be identified, it can be postulated that it exists in the Negative Valence Domain given that the symptoms relate predominantly to negative thoughts, mood, anxiety, and panic.

Obviously, as this study is exploratory, this example is extremely preliminary, and the exact findings associated with the Gulf War Cluster 5.2 in this study are not likely to be replicated to the point at which a construct can be defined. However, the overall philosophy underlying this study has been explicitly supported by Bruce Cuthbert, one of the initial authors of RDoCs, which is to use a sample of individuals with different diagnoses as per the DSM, or look at individuals within one DSM diagnoses for subtypes, and explore underlying differences in dimensions to inform a greater understanding of pathological mechanisms (Cuthbert, 2014).

Where this study differs from numerous others is in the concentration on identifying a trans-diagnostic population using only self-report symptom data. While in general this approach may be less compelling to basic researchers given the fact that the results are complex and difficult to interpret at the RDoC construct level, the symptom data is more relevant to patients seeking treatment and the clinicians who treat them, and to an extent is already being ascertained in clinical practice. It therefore appears prudent to include a broad range of self-report symptom ascertainment in future large psychiatric studies investigating identified or potential constructs. Continuing to focus only on the overall scoring of focused clinical inventories, as the majority of the RDoCs constructs currently do, essentially eliminates a valuable avenue of exploration that may ultimately produce relatively inexpensive methods of determining, through self-reported symptoms, which individuals should receive certain more expensive and invasive tests for diagnoses and treatment once the mechanisms for psychiatric distress become better understood and biomarkers are shown to have clinical utility.

There were many limitations in this study. Most notably, the number of individuals in the samples was smaller than optimal given the high number of variables used for clustering (Hastie et al., 2001). We did attempt dimensional reduction of variables with PCA as well as the mapping of symptoms to higher level symptom categories as described in Chapter 4 regarding workflow, but the clusters identified were not deemed as stable by internal validation metrics as those without feature reduction, so these were not used in the subsequent analyses. The large amount of symptoms used to create the symptom-level profiles also resulted in a lack of power to identify if any individual symptoms were significantly different across clusters. Additionally, there were too many variables used to externally validate the clusters (e.g., alcohol use, history of childhood trauma, neuropsychiatric measures, imaging, and neuroendocrine variables) for the vast majority of findings to meet criteria for statistical significance when applying standard multiple testing corrections. This study was conceived as an exploratory study, however, and its value in large part resides in its potential use as a source of information to aid researchers in evaluating prior findings in the literature, and in formulating hypotheses to test in future studies. Additionally, the fact that data was pooled from two separate studies was both a weakness and a strength. The Gulf War and MIRECC had different study designs, and subsequently different populations, and so while combining the samples created a large sample with greater power, there are shortcoming that can incur when samples were not ascertained in exactly the same manner (Ahrens & Pigeot, 2007).

This study also has limited generalizability, especially due to the fact that individuals with severe affective disorders and psychotic orders were excluded from the sample. Given our findings that clusters of individuals with both a greater intensity and breadth of symptoms had the most significant clinical and biological correlates, it will be of great interest to conduct a

similar study with the inclusion of individuals with psychotic symptoms that arguably are responsible for the most distress and impairment among patients who are mentally ill (Bowie & Harvey, 2006).

In conclusion, the results of this study support increasing the breadth of symptoms ascertained from individuals to use in defining more highly specified and homogeneous psychiatric phenotypes. This study also demonstrates that using machine-learning methods with these data may be effective in delineating these data-driven phenotypes. This approach is inherently dimensional, and allows for the inclusion of a larger breadth of symptoms than the DSM diagnostic algorithms. If the approach is taken to ascertain a wide variety of symptoms without the goal of identifying a DSM diagnosis, a greater number of symptoms will be ascertained and recorded, and can be used to facilitate data-driven psychiatric phenotypes.

There may be a level of psychopathology that can be identified through elevated symptoms across what has been classically perceived as different psychiatric domains, that identifies individuals with clinical and biological markers that diverge from those of individuals who experience less psychiatric distress. While symptoms may not be the only way to identify these individuals, it is worth looking at the utility of self-report symptoms as they require a relatively low level of resource to ascertain, especially now that many self-report symptoms can conceivably be gathered remotely from questionnaires over the internet. Additionally, the expanding use of EHRs and the ongoing development of Natural Language Processing algorithms may enable a large degree of automated ascertainment. Finally, the continued importance of using patient-reported symptom data to diagnose and treat individuals with psychiatric illness is necessary to form stable and healing relationships with patients, “to meet

the patient where they are at," and to address their concerns, which is the foundation of current initiatives to support the improvement of patient-centered outcomes in medicine.

References

Afifi, T. O., Mather, A., Boman, J., Fleisher, W., Enns, M. W., Macmillan, H., & Sareen, J. (2011). Childhood adversity and personality disorders: results from a nationally representative population-based study. *J Psychiatr Res, 45*(6), 814-822. doi:10.1016/j.jpsychires.2010.11.008

Agartz, I., Momenan, R., Rawlings, R. R., Kerich, M. J., & Hommer, D. W. (1999). Hippocampal volume in patients with alcohol dependence. *Arch Gen Psychiatry, 56*(4), 356-363.

Ahrens, W., & Pigeot, I. (2007). *Handbook of epidemiology with 165 figures and 180 tables* (pp. xi, 1639 p.). Retrieved from https://ucsf.idm.oclc.org/login?url=http://dx.doi.org/10.1007/b137839

Alsobrook, J. P., 2nd, & Pauls, D. L. (2002). A factor analysis of tic symptoms in Gilles de la Tourette's syndrome. *Am J Psychiatry, 159*(2), 291-296. doi:10.1176/appi.ajp.159.2.291

Alvarez, M. J., Roura, P., Oses, A., Foguet, Q., Sola, J., & Arrufat, F. X. (2011). Prevalence and clinical impact of childhood trauma in patients with severe mental disorders. *J Nerv Ment Dis, 199*(3), 156-161. doi:10.1097/NMD.0b013e31820c751c

American Psychiatric Association., & American Psychiatric Association. DSM-5 Task Force. *Diagnostic and statistical manual of mental disorders : DSM-5* (pp. 1 online resource (xliv, 947 pages)). Retrieved from https://ucsf.idm.oclc.org/login?url=http://uclibs.org/PID/261189

American Psychiatric Association., & American Psychiatric Association. Task Force on DSM-IV. (1994). *Diagnostic and statistical manual of mental disorders : DSM-IV* (4th ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association., & American Psychiatric Association. Work Group to Revise DSM-III. (1987). *Diagnostic and statistical manual of mental disorders : DSM-III-R* (3rd ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. Committee on Nomenclature and Statistics. (1952). *Mental disorders; diagnostic and statistical manual*. Washington,: American Psychiatric Association.

American Psychiatric Association. Committee on Nomenclature and Statistics. (1968). *Diagnostic and statistical manual of mental disorders* (2d ed.). Washington,: American Psychiatric Association.

Andreasen, N. C. (1995). Symptoms, signs, and diagnosis of schizophrenia. *Lancet, 346*(8973), 477-481.

Apfel, B. A., Ross, J., Hlavin, J., Meyerhoff, D. J., Metzler, T. J., Marmar, C. R., . . . Neylan, T. C. Hippocampal volume differences in gulf war veterans with current versus lifetime posttraumatic stress disorder symptoms. *Biol Psychiatry, 69*(6), 541-548. doi:S0006-3223(10)01013-9 [pii]
10.1016/j.biopsych.2010.09.044

Apfel, B. A., Ross, J., Hlavin, J., Meyerhoff, D. J., Metzler, T. J., Marmar, C. R., . . . Neylan, T. C. (2011). Hippocampal volume differences in Gulf War veterans with current versus

lifetime posttraumatic stress disorder symptoms. *Biol Psychiatry, 69*(6), 541-548. doi:10.1016/j.biopsych.2010.09.044

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Perez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition, 46*(1), 243-256. doi:10.1016/j.patcog.2012.07.021

Austin, M. P., Mitchell, P., & Goodwin, G. M. (2001). Cognitive deficits in depression: possible implications for functional neuropathology. *Br J Psychiatry, 178*, 200-206.

Baltieri, D. A., & Correa Filho, J. M. (2012). Role of two clusters of male alcoholics in treatment retention. *Eur Addict Res, 18*(4), 201-211. doi:10.1159/000336530

Bayati, H., Davoudi, H., & Fatemizadeh, E. (2008). A Heuristic Method for Finding the Optimal Number of Clusters with Application in Medical Data. *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-8*, 4684-4687. doi:Doi 10.1109/Iembs.2008.4650258

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Arch Gen Psychiatry, 4*, 561-571.

Bendall, S., Jackson, H. J., Hulbert, C. A., & McGorry, P. D. (2008). Childhood trauma and psychotic disorders: a systematic, critical review of the evidence. *Schizophr Bull, 34*(3), 568-579. doi:10.1093/schbul/sbm121

Bing, X., Ming-Guo, Q., Ye, Z., Jing-Na, Z., Min, L., Han, C., . . . Shao-Xiang, Z. (2013). Alterations in the cortical thickness and the amplitude of low-frequency fluctuation in patients with post-traumatic stress disorder. *Brain Res, 1490*, 225-232. doi:10.1016/j.brainres.2012.10.048

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a Clinician-Administered PTSD Scale. *J Trauma Stress, 8*(1), 75-90.

Bora, E., Yucel, M., & Pantelis, C. (2010). Cognitive impairment in schizophrenia and affective psychoses: implications for DSM-V criteria and beyond. *Schizophr Bull, 36*(1), 36-42. doi:10.1093/schbul/sbp094

Bowie, C. R., & Harvey, P. D. (2006). Cognitive deficits and functional outcome in schizophrenia. *Neuropsychiatr Dis Treat, 2*(4), 531-536.

Bremner, J. D., Narayan, M., Anderson, E. R., Staib, L. H., Miller, H. L., & Charney, D. S. (2000). Hippocampal volume reduction in major depression. *Am J Psychiatry, 157*(1), 115-118. doi:10.1176/ajp.157.1.115

Brewin, C. R., Andrews, B., & Valentine, J. D. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *J Consult Clin Psychol, 68*(5), 748-766.

Buckner, R. L., Head, D., Parker, J., Fotenos, A. F., Marcus, D., Morris, J. C., & Snyder, A. Z. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage, 23*(2), 724-738. doi:S1053811904003271 [pii]

10.1016/j.neuroimage.2004.06.018

Cardenas, V. A., Samuelson, K., Lenoci, M., Studholme, C., Neylan, T. C., Marmar, C. R., . . . Weiner, M. W. (2011). Changes in brain anatomy during the course of posttraumatic stress disorder. *Psychiatry Res, 193*(2), 93-100. doi:10.1016/j.pscychresns.2011.01.013

Carsten F. Dormann, Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, . . . Lautenbach, S. (2012). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*(36), 27-46. doi:10.1111/j.1600-0587.2012.07348.x

Chapman, D. P., Whitfield, C. L., Felitti, V. J., Dube, S. R., Edwards, V. J., & Anda, R. F. (2004). Adverse childhood experiences and the risk of depressive disorders in adulthood. *J Affect Disord, 82*(2), 217-225. doi:10.1016/j.jad.2003.12.013

Chen, J., Xu, T., Jing, J., & Chan, R. C. (2011). Alexithymia and emotional regulation: A cluster analytical approach. *BMC Psychiatry, 11*, 33. doi:10.1186/1471-244X-11-33

Christensen, G. E., Joshi, S. C., & Miller, M. I. (1997). Volumetric transformation of brain anatomy. *IEEE Trans Med Imaging, 16*(6), 864-877. doi:10.1109/42.650882

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *N Engl J Med, 372*(9), 793-795. doi:10.1056/NEJMp1500523

Connor, J. P., Haber, P. S., & Hall, W. D. (2016). Alcohol use disorders. *Lancet, 387*(10022), 988-998. doi:10.1016/S0140-6736(15)00122-1

Craddock, N., & Owen, M. J. (2010). The Kraepelinian dichotomy - going, going... but still not gone. *Br J Psychiatry, 196*(2), 92-95. doi:10.1192/bjp.bp.109.073429

Cross-Disorder Group of the Psychiatric Genomics, C., Smoller, J. W., Craddock, N., Kendler, K., Lee, P. H., Neale, B. M., . . . Sullivan, P. F. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet, 381*(9875), 1371-1379. doi:10.1016/S0140-6736(12)62129-1

Cuthbert, B. N. (2014). The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry, 13*(1), 28-35. doi:10.1002/wps.20087

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med, 11*, 126. doi:10.1186/1741-7015-11-126

Daniels, J. K., Lamke, J. P., Gaebler, M., Walter, H., & Scheel, M. (2013). White matter integrity and its relationship to PTSD and childhood trauma--a systematic review and meta-analysis. *Depress Anxiety, 30*(3), 207-216. doi:10.1002/da.22044

Dazzi, F., Shafer, A., & Lauriola, M. (2016). Meta-analysis of the Brief Psychiatric Rating Scale - Expanded (BPRS-E) structure and arguments for a new version. *J Psychiatr Res, 81*, 140-151. doi:10.1016/j.jpsychires.2016.07.001

de Kloet, C. S., Vermetten, E., Geuze, E., Kavelaars, A., Heijnen, C. J., & Westenberg, H. G. (2006). Assessment of HPA-axis function in posttraumatic stress disorder: pharmacological and non-pharmacological challenge tests, a review. *J Psychiatr Res, 40*(6), 550-567. doi:10.1016/j.jpsychires.2005.08.002

Delis, D. C., Freeland, J., Kramer, J. H., & Kaplan, E. (1988). Integrating clinical assessment with cognitive neuroscience: construct validation of the California Verbal Learning Test. *J Consult Clin Psychol, 56*(1), 123-130.

Derogatis, L. (1975). *Symptom Checklist-90-Revised*. Minneapolis, MN: NCS Assessments.

Drevets, W. C. (2000). Neuroimaging studies of mood disorders. *Biol Psychiatry, 48*(8), 813-829.

Driessen, M., Lange, W., Junghanns, K., & Wetterling, T. (2005). Proposal of a comprehensive clinical typology of alcohol withdrawal--a cluster analysis approach. *Alcohol Alcohol, 40*(4), 308-313. doi:10.1093/alcalc/agh167

Dube, S. R., Anda, R. F., Felitti, V. J., Edwards, V. J., & Croft, J. B. (2002). Adverse childhood experiences and personal alcohol abuse as an adult. *Addict Behav, 27*(5), 713-725.

Dube, S. R., Felitti, V. J., Dong, M., Chapman, D. P., Giles, W. H., & Anda, R. F. (2003). Childhood abuse, neglect, and household dysfunction and the risk of illicit drug use: the adverse childhood experiences study. *Pediatrics, 111*(3), 564-572.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* London: Chapman & Hall/CRC.

Everitt, B. (2011). *Cluster analysis* (5th ed.). Chichester, West Sussex, U.K.: Wiley.

Everitt, B., Ismail, K., David, A. S., & Wessely, S. (2002). Searching for a Gulf War syndrome using cluster analysis. *Psychol Med, 32*(8), 1371-1378.

Fals-Stewart, W., & Lucente, S. (1993). An MCMI cluster typology of obsessive-compulsives: a measure of personality characteristics and its relationship to treatment participation, compliance and outcome in behavior therapy. *J Psychiatr Res, 27*(2), 139-154.

Farmer, A. E., McGuffin, P., & Spitznagel, E. L. (1983). Heterogeneity in schizophrenia: a cluster-analytic approach. *Psychiatry Res, 8*(1), 1-12.

First, M. B., Spitzer, Robert L, Gibbon Miriam, and Williams, Janet B.W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P)*: New York State Psychiatric Institute,.

Franklin, C. L., & Zimmerman, M. (2001). Posttraumatic stress disorder and major depressive disorder: investigating the role of overlapping symptoms in diagnostic comorbidity. *J Nerv Ment Dis, 189*(8), 548-551.

Gabbard, G. O. (1990). *Psychodynamic psychiatry in clinical practice*. Washington, DC: American Psychiatric Press.

Gara, M. A., Silver, R. C., Escobar, J. I., Holman, A., & Waitzkin, H. (1998). A hierarchical classes analysis (HICLAS) of primary care patients with medically unexplained somatic symptoms. *Psychiatry Res, 81*(1), 77-86.

Garamszegi, L. Z. (2006). Comparing effect sizes across variables: generalization without the need for Bonferroni correction. *Behavioral Ecology, 17*(4), 682-687.

Gaziano, J. M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., . . . O'Leary, T. J. (2016). Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol, 70*, 214-223. doi:10.1016/j.jclinepi.2015.09.016

Geuze, E., Westenberg, H. G., Heinecke, A., de Kloet, C. S., Goebel, R., & Vermetten, E. (2008). Thinner prefrontal cortex in veterans with posttraumatic stress disorder. *Neuroimage, 41*(3), 675-681. doi:10.1016/j.neuroimage.2008.03.007

Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., . . . Etkin, A. (2015). Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry, 72*(4), 305-315. doi:10.1001/jamapsychiatry.2014.2206

Goodman, L. S., Gilman, A., Brunton, L. L., Lazo, J. S., & Parker, K. L. (2006). *Goodman & Gilman's the pharmacological basis of therapeutics* (11th ed.). New York: McGraw-Hill.

Goodyer, I. M. (2012). Mathematical models as an aid for improving the validity of descriptive psychiatry. *Br J Psychiatry, 201*(5), 335-336. doi:10.1192/bjp.bp.112.109538

Guclu, O., Senormanci, O., Aydin, E., Erkiran, M., & Kokturk, F. (2015). Phenomenological subtypes of mania and their relationships with substance use disorders. *J Affect Disord, 174*, 569-573. doi:10.1016/j.jad.2014.11.016

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal fo Machine Learning Research, 3*, 1157-1182.

Hales, R. E., Yudofsky, S. C., Gabbard, G. O., & American Psychiatric Publishing. *The American Psychiatric Publishing textbook of psychiatry* (pp. 1 online resource). Retrieved from https://ucsf.idm.oclc.org/login?url=http://dx.doi.org/10.1176/appi.books.9781585623402

Hall, W., Degenhardt, L., & Teesson, M. (2009). Understanding comorbidity between substance use, anxiety and affective disorders: broadening the research base. *Addict Behav, 34*(6-7), 526-530. doi:10.1016/j.addbeh.2009.03.010

Hamilton, M. (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry, 23*, 56-62.

Hand, D. J., & Heard, N. A. (2005). Finding groups in gene expression data. *J Biomed Biotechnol, 2005*(2), 215-225. doi:10.1155/JBB.2005.215

Hanson, J. L., Chung, M. K., Avants, B. B., Shirtcliff, E. A., Gee, J. C., Davidson, R. J., & Pollak, S. D. (2010). Early stress is associated with alterations in the orbitofrontal cortex: a tensor-based morphometry investigation of brain structure and behavioral risk. *J Neurosci, 30*(22), 7466-7472. doi:10.1523/JNEUROSCI.0859-10.2010

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations*. New York: Springer.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning : data mining, inference, and prediction* (2nd ed.). New York, NY: Springer.

Hay, P. J., Fairburn, C. G., & Doll, H. A. (1996). The classification of bulimic eating disorders: a community-based cluster analysis study. *Psychol Med, 26*(4), 801-812.

Heim, C., & Nemeroff, C. B. (2001). The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biol Psychiatry, 49*(12), 1023-1039.

Helmes, E., & Landmark, J. (2003). Subtypes of schizophrenia: a cluster analytic approach. *Can J Psychiatry, 48*(10), 702-708.

Hsu, Y. Y., Schuff, N., Du, A. T., Mark, K., Zhu, X., Hardin, D., & Weiner, M. W. (2002). Comparison of automated and manual MRI volumetry of hippocampus in normal aging and dementia. *J Magn Reson Imaging, 16*(3), 305-310. doi:10.1002/jmri.10163

Hubert, L. A., P. (1985). Comparing partitions. *Journal of Classification, 2*.

Hyman, S. E. (2010). The diagnosis of mental disorders: the problem of reification. *Annu Rev Clin Psychol, 6*, 155-179. doi:10.1146/annurev.clinpsy.3.022806.091532

Insel, T. R. (2014). The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. *Am J Psychiatry, 171*(4), 395-397. doi:10.1176/appi.ajp.2014.14020138

Iovieno, N., Tedeschini, E., Bentley, K. H., Evins, A. E., & Papakostas, G. I. (2011). Antidepressants for major depressive disorder and dysthymic disorder in patients with comorbid alcohol use disorders: a meta-analysis of placebo-controlled randomized trials. *J Clin Psychiatry, 72*(8), 1144-1151. doi:10.4088/JCP.10m06217

Iwao, K., Matoba, R., Ueno, N., Ando, A., Miyoshi, Y., Matsubara, K., . . . Kato, K. (2002). Molecular classification of primary breast tumors possessing distinct prognostic properties. *Hum Mol Genet, 11*(2), 199-206.

Jablensky, A. (1999). The nature of psychiatric classification: issues beyond ICD-10 and DSM-IV. *Aust N Z J Psychiatry, 33*(2), 137-144.

Jablensky, A., & Waters, F. (2014). RDoC: a roadmap to pathogenesis? *World Psychiatry, 13*(1), 43-44. doi:10.1002/wps.20100

Jackson, C. T., Allen, G., Essock, S. M., Foster, M. J., Lanzara, C. B., Felton, C. J., & Donahue, S. A. (2006). Clusters of event reactions among recipients of Project Liberty mental health counseling. *Psychiatr Serv, 57*(9), 1271-1276. doi:10.1176/appi.ps.57.9.1271

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning : with applications in R*. New York: Springer.

Karl, A., Schaefer, M., Malta, L. S., Dorfel, D., Rohleder, N., & Werner, A. (2006). A meta-analysis of structural brain abnormalities in PTSD. *Neurosci Biobehav Rev, 30*(7), 1004-1031. doi:10.1016/j.neubiorev.2006.03.004

Kirschbaum, C., & Hellhammer, D. H. (1989). Salivary cortisol in psychobiological research: an overview. *Neuropsychobiology, 22*(3), 150-169. doi:118611

Lee, J., Im, S. J., Lee, S. G., Stadlin, A., Son, J. W., Shin, C. J., . . . Kim, S. (2016). Volume of hippocampal subfields in patients with alcohol dependence. *Psychiatry Res, 258*, 16-22. doi:10.1016/j.pscychresns.2016.10.009

Lev-Ran, S., Balchand, K., Lefebvre, L., Araki, K. F., & Le Foll, B. (2012). Pharmacotherapy of alcohol use disorders and concurrent psychiatric disorders: a review. *Can J Psychiatry, 57*(6), 342-349.

Li, L., Wu, M., Liao, Y., Ouyang, L., Du, M., Lei, D., . . . Gong, Q. (2014). Grey matter reduction associated with posttraumatic stress disorder and traumatic stress. *Neurosci Biobehav Rev, 43*, 163-172. doi:10.1016/j.neubiorev.2014.04.003

Lisboa, P. J., Etchells, T. A., Jarman, I. H., & Chambers, S. J. (2013). Finding reproducible cluster partitions for the k-means algorithm. *BMC Bioinformatics, 14 Suppl 1*, S8. doi:10.1186/1471-2105-14-S1-S8

Lochner, C., du Toit, P. L., Zungu-Dirwayi, N., Marais, A., van Kradenburg, J., Seedat, S., . . . Stein, D. J. (2002). Childhood trauma in obsessive-compulsive disorder, trichotillomania, and controls. *Depress Anxiety, 15*(2), 66-68.

Maj, M. (2005). "Psychiatric comorbidity": an artefact of current diagnostic systems? *Br J Psychiatry, 186*, 182-184. doi:10.1192/bjp.186.3.182

Maj, M. (2014). Keeping an open attitude towards the RDoC project. *World Psychiatry, 13*(1), 1-3. doi:10.1002/wps.20111

Mathews, C. A., Jang, K. L., Herrera, L. D., Lowe, T. L., Budman, C. L., Erenberg, G., . . . Reus, V. I. (2007). Tic symptom profiles in subjects with Tourette Syndrome from two genetically isolated populations. *Biol Psychiatry, 61*(3), 292-300. doi:10.1016/j.biopsych.2006.02.009

McCoy, T. H., Castro, V. M., Rosenfield, H. R., Cagan, A., Kohane, I. S., & Perlis, R. H. (2015). A clinical perspective on the relevance of research domain criteria in electronic health records. *Am J Psychiatry, 172*(4), 316-320. doi:10.1176/appi.ajp.2014.14091177

McCrea, R. (2013). *Rethinking the nature of mental disorder:a latent structure approach to data from three national psychiatric morbidity surveys.* (Doctor of Philosophy), University College London, London.

McTeague, L. M., Goodkind, M. S., & Etkin, A. (2016). Transdiagnostic impairment of cognitive control in mental illness. *J Psychiatr Res, 83*, 37-46. doi:10.1016/j.jpsychires.2016.08.001

Meyer, H. (2011). Coding complexity: US Health Care gets ready for the coming Of ICD-10. *Health Aff (Millwood), 30*(5), 968-974. doi:10.1377/hlthaff.2011.0319

Murthy, P., & Chand, P. (2012). Treatment of dual diagnosis disorders. *Curr Opin Psychiatry, 25*(3), 194-200. doi:10.1097/YCO.0b013e328351a3e0

Muthen, B. (2001). LCA and Cluster Analysis. [07/01/2016]. http://www.statmodel.com/.

Muthen, B., Asparouhov, T., & Rebollo, I. (2006). Advances in behavioral genetics modeling using Mplus: applications of factor mixture modeling to twin data. *Twin Res Hum Genet, 9*(3), 313-324. doi:10.1375/183242706777591317

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide: Statistical Analysis with Latent Variables* (7th ed.). Los Angeles, CA: Muthén & Muthén.

National Research Council (U.S.). Committee on A Framework for Developing a New Taxonomy of Disease. (2011). *Toward precision medicine : building a knowledge network for biomedical research and a new taxonomy of disease*. Washington, D.C.: National Academies Press.

Neylan, T. C., Brunet, A., Pole, N., Best, S. R., Metzler, T. J., Yehuda, R., & Marmar, C. R. (2005). PTSD symptoms predict waking salivary cortisol levels in police officers. *Psychoneuroendocrinology, 30*(4), 373-381. doi:10.1016/j.psyneuen.2004.10.005

Oquendo, M. A., Echavarria, G., Galfalvy, H. C., Grunebaum, M. F., Burke, A., Barrera, A., . . . John Mann, J. (2003). Lower cortisol levels in depressed patients with comorbid post-traumatic stress disorder. *Neuropsychopharmacology, 28*(3), 591-598. doi:10.1038/sj.npp.1300050

Ortoleva Bucher, C., Dubuc, N., von Gunten, A., Trottier, L., & Morin, D. (2016). Development and validation of clinical profiles of patients hospitalized due to behavioral and psychological symptoms of dementia. *BMC Psychiatry, 16*, 261. doi:10.1186/s12888-016-0966-7

Ozer, E. J., Best, S. R., Lipsey, T. L., & Weiss, D. S. (2003). Predictors of posttraumatic stress disorder and symptoms in adults: a meta-analysis. *Psychol Bull, 129*(1), 52-73.

Paykel, E. S. (1971). Classification of depressed patients: a cluster analysis derived grouping. *Br J Psychiatry, 118*(544), 275-288.

Perlis, R. H., Iosifescu, D. V., Castro, V. M., Murphy, S. N., Gainer, V. S., Minnier, J., . . . Smoller, J. W. (2012). Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. *Psychol Med, 42*(1), 41-50. doi:10.1017/S0033291711000997

Pettinati, H. M., O'Brien, C. P., & Dundon, W. D. (2013). Current status of co-occurring mood and substance use disorders: a new therapeutic target. *Am J Psychiatry, 170*(1), 23-30. doi:10.1176/appi.ajp.2012.12010112

Pilowsky, I., Levine, S., & Boulton, D. M. (1969). The classification of depression by numerical taxonomy. *Br J Psychiatry, 115*(525), 937-945.

Pitman, R. K., Rasmusson, A. M., Koenen, K. C., Shin, L. M., Orr, S. P., Gilbertson, M. W., . . . Liberzon, I. (2012). Biological studies of post-traumatic stress disorder. *Nat Rev Neurosci, 13*(11), 769-787. doi:10.1038/nrn3339

Regier, D. A., Farmer, M. E., Rae, D. S., Locke, B. Z., Keith, S. J., Judd, L. L., & Goodwin, F. K. (1990). Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. *JAMA, 264*(19), 2511-2518.

Renwick, B., Musiat, P., Lose, A., DeJong, H., Broadbent, H., Kenyon, M., . . . Schmidt, U. (2015). Neuro- and social-cognitive clustering highlights distinct profiles in adults with anorexia nervosa. *Int J Eat Disord, 48*(1), 26-34. doi:10.1002/eat.22366

Reser, M. P., Allott, K. A., Killackey, E., Farhall, J., & Cotton, S. M. (2015). Exploring cognitive heterogeneity in first-episode psychosis: What cluster analysis can reveal. *Psychiatry Res, 229*(3), 819-827. doi:10.1016/j.psychres.2015.07.084

Robertson, H. A., Lam, R. W., Stewart, J. N., Yatham, L. N., Tam, E. M., & Zis, A. P. (1996). Atypical depressive symptoms and clusters in unipolar and bipolar depression. *Acta Psychiatr Scand, 94*(6), 421-427.

Rosen, G. M., Lilienfeld, S. O., Frueh, B. C., McHugh, P. R., & Spitzer, R. L. (2010). Reflections on PTSD's future in DSM-V. *Br J Psychiatry, 197*(5), 343-344. doi:10.1192/bjp.bp.110.079699

Ross, J., Neylan, T., Weiner, M., Chao, L., Samuelson, K., & Sim, I. (2015). Towards Constructing a New Taxonomy for Psychiatry Using Self-reported Symptoms. *Stud Health Technol Inform, 216*, 736-740.

Samuelson, K. W. (2011). Post-traumatic stress disorder and declarative memory functioning: a review. *Dialogues Clin Neurosci, 13*(3), 346-351.

Samuelson, K. W., Neylan, T. C., Metzler, T. J., Lenoci, M., Rothlind, J., Henn-Haase, C., . . . Marmar, C. R. (2006). Neuropsychological functioning in posttraumatic stress disorder and alcohol abuse. *Neuropsychology, 20*(6), 716-726. doi:2006-20657-010 [pii]
10.1037/0894-4105.20.6.716

Savic, D., Knezevic, G., Damjanovic, S., Spiric, Z., & Matic, G. (2012). The role of personality and traumatic events in cortisol levels--where does PTSD fit in? *Psychoneuroendocrinology, 37*(7), 937-947. doi:10.1016/j.psyneuen.2011.11.001

Schilling, E. A., Aseltine, R. H., Jr., & Gore, S. (2007). Adverse childhood experiences and mental health in young adults: a longitudinal survey. *BMC Public Health, 7*, 30. doi:10.1186/1471-2458-7-30

Schneeberger, A. R., Huber, C. G., Seixas, A., Muenzenmaier, K. H., Lang, U. E., Castille, D., . . . Link, B. G. (2017). Alcohol consumption and the use of health care services in people with severe mental illness and stressful childhood experiences. *J Addict Dis*, 0. doi:10.1080/10550887.2017.1280311

Schuff, N., Neylan, T. C., Fox-Bosetti, S., Lenoci, M., Samuelson, K. W., Studholme, C., . . . Weiner, M. W. (2008). Abnormal N-acetylaspartate in hippocampus and anterior cingulate in posttraumatic stress disorder. *Psychiatry Res, 162*(2), 147-157. doi:S0925-4927(07)00099-6 [pii]
10.1016/j.pscychresns.2007.04.011

Segal, D. L., Hersen, M., & Van Hasselt, V. B. (1994). Reliability of the Structured Clinical Interview for DSM-III-R: an evaluative review. *Compr Psychiatry, 35*(4), 316-327.

Sharp, C., Fowler, J. C., Salas, R., Nielsen, D., Allen, J., Oldham, J., . . . Fonagy, P. (2016). Operationalizing NIMH Research Domain Criteria (RDoC) in naturalistic clinical settings. *Bull Menninger Clin, 80*(3), 187-212. doi:10.1521/bumc.2016.80.3.187

Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *Am Psychol, 65*(2), 98-109. doi:10.1037/a0018378

Shiino, A., Matsuda, M., Morikawa, S., Inubushi, T., Akiguchi, I., & Handa, J. (1993). Proton magnetic resonance spectroscopy with dementia. *Surg Neurol, 39*(2), 143-147.

Silverman, M. H. K., R.F. (2014). Borderline Personality Disorder and DSM-5: New Directions and hopes for the future. In C. S. J.Tackett (Ed.), *Handbook of borderline personality disorder in children and adolescents* (pp. 433-450). New York, New York: Springer.

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. *Front Psychol, 6*, 328. doi:10.3389/fpsyg.2015.00328

Soares, J. C., & Mann, J. J. (1997). The anatomy of mood disorders--review of structural neuroimaging studies. *Biol Psychiatry, 41*(1), 86-106.

Sullivan, P. F., Kessler, R. C., & Kendler, K. S. (1998). Latent class analysis of lifetime depressive symptoms in the national comorbidity survey. *Am J Psychiatry, 155*(10), 1398-1406. doi:10.1176/ajp.155.10.1398

Team, R. D. C. (2011). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.r-project.org/

Trivedi, J. K. (2006). Cognitive deficits in psychiatric disorders: Current status. *Indian J Psychiatry, 48*(1), 10-20. doi:10.4103/0019-5545.31613

Tsuang, M. T., Tohen, M., & Jones, P. B. (2011). *RC Kessler*

The National Comorbidity Survey (NCS) and its extensions

Textbook of psychiatric epidemiology (3rd ed.). Chichester, West Sussex ; Hoboken, NJ: Wiley-Blackwell.

Turner, J. A. (2014). The rise of large-scale imaging studies in psychiatry. *Gigascience, 3*, 29. doi:10.1186/2047-217X-3-29

Uebersax, J. (1999). Probit Latent Class Analysis with Dichotomous or Ordered Category Measures: Conditional Independence/Dependence Models. *Applied Psychological Measurement, 23*(4), 283-297.

Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2005). Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Arch Gen Psychiatry, 62*(6), 629-640. doi:10.1001/archpsyc.62.6.629

Wardlaw, J. M., Valdes Hernandez, M. C., & Munoz-Maniega, S. (2015). What are white matter hyperintensities made of? Relevance to vascular cognitive impairment. *J Am Heart Assoc, 4*(6), 001140. doi:10.1161/JAHA.114.001140

Welch, K. A., Carson, A., & Lawrie, S. M. (2013). Brain structure in adolescents and young adults with alcohol problems: systematic review of imaging studies. *Alcohol Alcohol, 48*(4), 433-444. doi:10.1093/alcalc/agt037

Wessman, J., Paunio, T., Tuulio-Henriksson, A., Koivisto, M., Partonen, T., Suvisaari, J., . . . Peltonen, L. (2009). Mixture model clustering of phenotype features reveals evidence for association of DTNBP1 to a specific subtype of schizophrenia. *Biol Psychiatry, 66*(11), 990-996. doi:10.1016/j.biopsych.2009.05.034

Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck, MacCallum, R. C. (Ed.), *Factor analysis at 100: Historical developments and future directions*. Mahwah, HJ: Erlbaum.

Williams, G. W., Barton, G. M., White, A. A., & Won, H. (1976). Cluster analysis applied to symptom ratings of psychiatric patients: an evaluation of its predictive ability. *Br J Psychiatry, 129*, 178-185.

Wolf, E. J., Sadeh, N., Leritz, E. C., Logue, M. W., Stoop, T. B., McGlinchey, R., . . . Miller, M. W. (2016). Posttraumatic Stress Disorder as a Catalyst for the Association Between Metabolic Syndrome and Reduced Cortical Thickness. *Biol Psychiatry, 80*(5), 363-371. doi:10.1016/j.biopsych.2015.11.023

Woodward, S. H., Kaloupek, D. G., Streeter, C. C., Kimble, M. O., Reiss, A. L., Eliez, S., . . . Arsenault, N. J. (2007). Brain, skull, and cerebrospinal fluid volumes in adult posttraumatic stress disorder. *J Trauma Stress, 20*(5), 763-774. doi:10.1002/jts.20241

Woodward, S. H., Schaer, M., Kaloupek, D. G., Cediel, L., & Eliez, S. (2009). Smaller global and regional cortical volume in combat-related posttraumatic stress disorder. *Arch Gen Psychiatry, 66*(12), 1373-1382. doi:66/12/1373 [pii]

10.1001/archgenpsychiatry.2009.160

Yehuda, R., Boisoneau, D., Lowy, M. T., & Giller, E. L., Jr. (1995). Dose-response changes in plasma cortisol and lymphocyte glucocorticoid receptors following dexamethasone administration in combat veterans with and without posttraumatic stress disorder. *Arch Gen Psychiatry, 52*(7), 583-593.

Yehuda, R., Boisoneau, D., Mason, J. W., & Giller, E. L. (1993). Glucocorticoid receptor number and cortisol excretion in mood, anxiety, and psychotic disorders. *Biol Psychiatry, 34*(1-2), 18-25. doi:0006-3223(93)90252-9 [pii]

APPENDIX A: Mapping of Neuropsychiatric Variables to Neuropsychiatric Tests

Neuropsychiatric Variable Name	**Original Neuropsychiatric Test**
Executive Functioning 1	ShortFreeCorrect
Executive Functioning 2	ShortCuedCorrect
Intrusions 1	TotalIntrusions
Intrusions 2	FreeRecallIntrusions
Intrusions 3	CuedRecallIntrusions
Learning 1	ImmediateRecallCorrectTrial1
Learning 2	ImmediateRecallCorrectTrial5
Learning 3	ImmediateRecallCorrectTotal
Learning 4	ImmediateRecallCorrectTrialB
Memory 1	LongFreeCorrect
Memory 2	LongCuedCorrect
Memory 3	logical_memory_immed_recall_total
Memory 4	logical_memory_delay_total
Performance IQ 1	wais3_digit_symbol_coding
Performance IQ 2	wais3_block_design
Repetitions	TotalRepetitions
Verbal Fluency 1	cowat_fas
Verbal Fluency 2	cowat_animals
Verbal IQ 1	wais3_vocabulary
Verbal IQ 2	wais3_similarities
Verbal IQ 3	wais3_digit_span
Verbal IQ 4	wais3_information

APPENDIX B: Mapping of American Psychiatric Association Symptom (APA) List to the Higher Level (HL) Symptom List

APA Symptom List	Higher Level Symptom List
Absent or Abnormal Social Play in Infancy or Childhood	
Aggressive Behavior	Behavior
Anhedonia	Neurovegetative
Antisocial Behavior	
Anxiety	Mood
Apathy	
Appetite Disturbance	Neurovegetative
Avoidance Behavior	Behavior
Binge Eating	Behavior
Blunted/Flat/Constricted Affect	
Catatonia	
Cross-Dressing	
Decrease in Energy or Fatigue	Neurovegetative
Delusions	Psychotic
Depersonalization or Derealization	
Depressed Mood	Mood
Disorganized Speech/Incoherence	
Distractibility	
Elevated Mood/Euphoria	
Feigning of Symptoms	
Flight of Ideas	
Grandiosity	
Grossly Disorganized Behavior	
Hallucinations	
Hypersomnia	
Impaired Abstract Thinking	
Impaired Judgment	
Inability to Maintain Attention/Poor Concentration	Neurovegetative
Increase in Social/Occupational/Sexual Activity	
Indecisiveness	Neurovegetative
Indifferent to Feelings of Others	
Indiscriminate Socializing	
Insomnia	Neurovegetative
Interpersonal Exploitativeness	
Irritability	Mood
Labile Affect	Mood
Memory Impairment	Neurovegetative
Paranoid Ideation	Psychotic
Persistent Identity Disturbance	
Physical Complaint Without General Medical Explanation	Somatic
Pressured Speech	
Psychomotor Agitation/Restlessness	Neurovegetative

Psychomotor Retardation	Neurovegetative
Repeated Lying	
Restricted Travel Away from Home	Behavior
Self-Induced Vomiting	
Self-Mutilating Behavior	
Sexual Dysfunction	Neurovegetative
Social Isolation	
Speech Difficulties	
Suicidal Ideation/Suicide Attempt	Mood
Weight Gain	
Weight Loss	Neurovegetative

APPENDIX C1: Mapping of Clinician's Assessment for PTSD (CAPS) to APA and HL Symptom Lists

CAPS Item	CAPS Variable	Label on CAPS Plots	APA Symptom Mapping	Higher Level Symptom Mapping and Symptom Number
Event	Criterion A event	Crit. A Event 1		
CAPSA1a	Life threat	Response to Event 1		
CAPS1bs	Life threat to self	Response to Event 2		
CAPS1bso	Life threat to other	Response to Event 3		
CAPSA2a	Threat to physical integrity	Response to Event 4		
CAPSA2bs	Threat to self (physical integrity)	Response to Event 5		
CAPSA2bso	Threat to others (physical integrity)	Response to Event 6		
CAPSA3a	Intense helplessness/fear/horror	Response to Event 7		
CAPSA3bd	Intense feelings during event	Response to Event 8		
CAPSA3ba	Intense feelings after event	Response to Event 9		
CAPSA4a	Time since event, years	Response to Event 10		
CAPSA5	Criterion A met	Response to Event 11		
CAPSB1a	B1-reexperiencing symptoms	Intrusive Symptoms 1	Inability to Concentrate	Neurovegetative 1
CAPSB1b		Intrusive Symptoms 2	Inability to Concentrate	Neurovegetative 2
CAPSB2a	B2-distressing dreams	Intrusive Symptoms 3	Insomnia	Neurovegetative 3
CAPSB2b		Intrusive Symptoms 4	Insomnia	Mood 35
CAPSB3a	B3-act/feel as if event recurring	Intrusive Symptoms 5	Inability to Concentrate	Neurovegetative 5
CAPSB3b		Intrusive Symptoms 6	Inability to Concentrate	Neurovegetative 6
CAPSB4a	B4-psycholgical distress w/exposure to cues	Intrusive Symptoms 7	Anxiety	Mood 1
CAPSB4b		Intrusive Symptoms 8	Anxiety	Mood 2
CAPSB5a	B5-physiological reaction w/exposure to cues	Intrusive Symptoms 9	Physical Complaint Without General Medical Explanation	Neurovegetative 7
CAPSB5b		Intrusive Symptoms 10	Physical Complaint Without General Medical Explanation	Somatic 2
CAPSC1a	C1-avoidance of thoughts, feelings, conversations	Avoidant Symptoms 1	Avoidance Behavior	Behavior 1
CAPSC1b		Avoidant Symptoms 2	Avoidance Behavior	Behavior 2
CAPSC2a	C2-avoidance of activities,	Avoidant Symptoms 3	Avoidance	Behavior 3

	places, people		Behavior	
CAPSC2b		Avoidant Symptoms 4	Avoidance Behavior	Behavior 4
CAPSC3a	C3-inability to recall important aspects of trauma	Avoidant Symptoms 5	Memory Impairment	Psychotic 12
CAPSC3b		Avoidant Symptoms 6	Memory Impairment	Neurovegetative 8
CAPSC4a	C4-diminished participation in activities	Avoidant Symptoms 7	Anhedonia	Mood 3
CAPSC4b		Avoidant Symptoms 8	Anhedonia	Mood 4
CAPSC5a	C5-detachment or estrangement	Avoidant Symptoms 9	Interpersonal Sensitivity	Interpersonal Sensitivity 4
CAPSC5b		Avoidant Symptoms 10	Depressed Mood	Mood 6
CAPSC6a	C6-restricted range of affect	Avoidant Symptoms 11	Depressed Mood	Mood 7
CAPSC6b		Avoidant Symptoms 12	Depressed Mood	Mood 8
CAPSC7a	C7-sense of foreshortened future	Avoidant Symptoms 13	Depressed Mood	Mood 9
CAPSC7b		Avoidant Symptoms 14	Depressed Mood	Mood 10
CAPSD1a	D1-difficulty falling/staying asleep	Arousal Symptoms 1	Insomnia	Mood 13
CAPSD1b		Arousal Symptoms 2	Insomnia	Neurovegetative 10
CAPSD2a	D2-irritability or outbursts of anger	Arousal Symptoms 3	Irritability	Mood 11
CAPSD2b		Arousal Symptoms 4	Irritability	Mood 12
CAPSD3a	D3-difficulty concentrating	Arousal Symptoms 5	Inability to Concentrate	Neurovegetative 11
CAPSD3b		Arousal Symptoms 6	Inability to Concentrate	Neurovegetative 12
CAPSD4a	D4-hypervigilance	Arousal Symptoms 7	Anxiety	Mood 43
CAPSD4b		Arousal Symptoms 8	Inability to Concentrate	Mood 14
CAPSD5a	D5-exaggerated startle response	Arousal Symptoms 9	Anxiety	Mood 15
CAPSD5b		Arousal Symptoms 10	Anxiety	Mood 16
CAPSEa	Duration of symptoms	Functional Impairment 1		
CAPSF1	Subjective distress	Functional Impairment 2		
CAPSF2	Impairment in social functioning	Functional Impairment 3		
CAPSF3	Impairment in occupational functioning	Functional Impairment 4		

APPENDIX C2: Mapping of Beck's Depression Inventory (BDI) to APA and HL Symptom Lists

Item #	Item	Graph Label	Map to APA Symptoms	Map to HL Symptoms
1	Sad	Sad	Depressed	Mood
2	Hopeless	Hopeless	Depressed	Mood
3	Feeling like a failure	Failure	Depressed	Mood
4	Loss of interest	Amotivation	Anhedonia	Neurovegetative
5	Guilty	Guilt	Depressed	Mood
6	Feelings of being punished	Punished	Depressed	Mood
7	Self-hatred	Self-Hate	Depressed	Mood
8	Self-blame	Self-Blame	Depressed	Mood
9	Suicidal ideation	Suicidality	Suicidal Ideation or Attempt	Mood
10	Crying	Crying	Depressed	Mood
11	Irritable	Irritable	Irritability	Mood
12	No satisfaction	Anhedonia	Anhedonia	Mood
13	Trouble with decisions	Indecisive	Indecisiveness	Behavior
14	Feeing ugly	Ugly	Depressed	Mood
15	Difficulty with Concentration	Poor Concentration	Poor Attention/Con-centration	Neurovegetative
16	Early morning wakening	Insomnia	Insomnia	Neurovegetative
17	Fatigue	Fatigue	Decrease in Energy or Fatigue	Neurovegetative
18	Appetite disturbance	Loss of Appetite	Appetite Disturbance	Neurovegetative
19	a. Weight loss	Weight Loss	Weight Loss	Neurovegetative
19	b. Trying to diet	Trying to Diet	N/A	N/A
20	Somatically focused	Somatically Focused	Physical Complaint Without GME	Somatic
21	Loss of libido	Loss of Libido	Physical Complaint Without GME	Somatic

APPENDIX C3: Mapping of Symptom Checklist 90 (SCL-90) to APA and HL Symptom Lists

SCL Question Number	Graph X-Axis Label	SCL Inventory Question	APA Symptom Mapping	Higher Level Symptom Mapping
81	HOS_81	Shouting/throwing things	Aggressive Behavior	Behavior 12
32	DEP_32	No interest in things	Anhedonia	Neurovegetative 9
2	ANX_2	Anxiety	Anxiety	Mood 30
23	ANX_23	Acutely scared for no reason	Anxiety	Mood 37
33	ANX_33	Fearful	Anxiety	Interpersonal Sensitivity 5
72	ANX_72	Panic spells	Anxiety	Mood 54
80	ANX_80	Ominous feelings	Anxiety	Mood 58
86	ANX_86	Frightening thoughts/images	Anxiety	Mood 60
31	DEP_31	Overworrying	Anxiety	Mood 42
10	OCD_10	Worried about carelessness	Anxiety	Mood 31
45	OCD_45	Double-check everything	Anxiety	Mood 47
65	OCD_65	Compulsions	Anxiety	Mood 52
47	PHOB_47	Afraid to travel	Anxiety	Mood 48
75	PHOB_75	Nervous alone	Anxiety	Mood 56
82	PHOB_82	Fear of fainting in public	Anxiety	Mood 59
17	ANX_17	Trembling	Anxiety	Mood 34
39	ANX_39	Heart pounding	Anxiety	Mood 46
22	DEP_22	Feeling trapped	Anxiety	Mood 36
38	OCD_38	Complete tasks slowly b/c need perfection	Anxiety	Mood 45
19	EAT_19	Poor appetite	Appetite Disturbance	Neurovegetative 22
60	EAT_60	Overeating	Binge Eating	Behavior 10
71	DEP_71	Everything is an effort	Decrease in Energy or Fatigue	Neurovegetative 31
14	DEP_14	Low energy	Decrease in Energy or Fatigue	Neurovegetative 21
90	PSY_90	Something wrong with your mind	Delusions	Psychotic 14
68	PI_68	Have beliefs others don't have	Delusions	Psychotic 8
62	PSY_62	Have thoughts that are not your own	Delusions	Psychotic 7
7	PSY_7	Others control thoughts	Delusions	Psychotic 1
84	PSY_84	Thoughts of sex that disturb you	Delusions	Psychotic 11
85	PSY_85	Feeling you should be punished	Delusions	Somatic 1
87	PSY_87	Body dysmorphism	Delusions	Psychotic 13
59	D2_59	Thoughts of death	Depressed Mood	Behavior 6
89	D2_89	Guilt	Depressed Mood	Mood 62
26	DEP_26	Self-blame	Depressed Mood	Mood 39
29	DEP_29	Feeling lonely	Depressed Mood	Mood 40
30	DEP_30	Feeling blue	Depressed Mood	Mood 41
54	DEP_54	Hopeless about future	Depressed Mood	Mood 49
79	DEP_79	Worthlessness	Depressed Mood	Mood 57
88	PSY_88	Feeling separate from others	Depressed Mood	Neurovegetative 32

16	PSY_16	Auditory hallucinations	Hallucinations	Psychotic 3
28	OCD_28	Feeling blocked	Inability to Concentrate	Neurovegetative 23
3	OCD_3	Intrusive thoughts	Inability to Concentrate	Neurovegetative 18
51	OCD_51	Mind going blank	Inability to Concentrate	Neurovegetative 25
55	OCD_55	Trouble concentrating	Inability to Concentrate	Neurovegetative 26
46	OCD_46	Indecisiveness	Indecisiveness	Behavior 8
44	SLEEP_44	Can't fall asleep	Insomnia	Neurovegetative 24
64	SLEEP_64	Early waking	Insomnia	Neurovegetative 29
66	SLEEP_66	Restless sleep	Insomnia	Neurovegetative 30
21	INT_21	Uneasy with opposite sex	Interpersonal Sensitivity	Interpersonal Sensitivity 2
34	INT_34	Feelings easily hurt	Interpersonal Sensitivity	Interpersonal Sensitivity 3
36	INT_36		Interpersonal Sensitivity	Mood 5
37	INT_37	Feeling disliked	Interpersonal Sensitivity	Mood 44
41	INT_41	Feeling Inferior	Interpersonal Sensitivity	Interpersonal Sensitivity 6
6	INT_6	Critical of others	Interpersonal Sensitivity	Interpersonal Sensitivity 1
61	INT_61	Uneasy with people watching you	Interpersonal Sensitivity	Interpersonal Sensitivity 7
69	INT_69	Very self-conscious	Interpersonal Sensitivity	Interpersonal Sensitivity 8
73	INT_73	Uneasy eating in public	Interpersonal Sensitivity	Interpersonal Sensitivity 9
77	PSY_77	Lonely with people	Interpersonal Sensitivity	Interpersonal Sensitivity 10
11	HOS_11	Easily irritated	Irritability	Mood 32
63	HOS_63	Urge to harm someone	Irritability	Mood 51
67	HOS_67	Urge to break things	Irritability	Mood 50
74	HOS_74	Frequent arguments	Irritability	Mood 55
20	DEP_20	Crying easily	Labile Affect	Neurovegetative 4
24	HOS_24	Temper outbursts	Labile Affect	Mood 38
9	OCD_9	Can't remember things	Memory Impairment	Neurovegetative 20
18	PI_18	Can't trust others	Paranoid Ideation	Psychotic 4
43	PI_43	Others watching you	Paranoid Ideation	Psychotic 6
76	PI_76	Not getting proper credit	Paranoid Ideation	Psychotic 9
8	PI_8	Others make your problems	Paranoid Ideation	Psychotic 2
83	PI_83	Feel people are trying to take advantage of you	Paranoid Ideation	Psychotic 10
35	PSY_35	Others know your thoughts	Paranoid Ideation	Psychotic 5
1	SOM_1	Headaches	Physical Complaint Without General Medical Explanation	Somatic 5
12	SOM_12	Chest pains	Physical Complaint Without General Medical Explanation	Somatic 7
27	SOM_27	Lower back pain	Physical Complaint Without General Medical Explanation	Somatic 8
4	SOM_4	Faintness/dizziness	Physical Complaint Without General Medical Explanation	Somatic 6
40	SOM_40	Nausea/GI upset	Physical Complaint	Somatic 9

			Without General Medical Explanation	
42	SOM_42	Muscle soreness	Physical Complaint Without General Medical Explanation	Somatic 10
48	SOM_48	Shortness of breath	Physical Complaint Without General Medical Explanation	Somatic 11
49	SOM_49	Hot or cold spells	Physical Complaint Without General Medical Explanation	Somatic 12
52	SOM_52	Numbness/tingling	Physical Complaint Without General Medical Explanation	Somatic 13
53	SOM_53	Lump in throat	Physical Complaint Without General Medical Explanation	Somatic 14
56	SOM_56	Weakness in body	Physical Complaint Without General Medical Explanation	Somatic 15
78	ANX_78	Restless	Psychomotor Agitation/Restlessness	Mood 61
57	ANX_57	Feeling keyed up	Psychomotor Agitation/Restlessness	Neurovegetative 27
58	SOM_58	Heavy limbs	Psychomotor Retardation	Neurovegetative 28
13	PHOB_13	Agoraphobia	Restricted Travel Away From Home	Mood 53
25	PHOB_25	Fear of leaving house	Restricted Travel Away From Home	Behavior 7
50	PHOB_50	Avoiding things that frighten you	Restricted Travel Away From Home	Behavior 9
70	PHOB_70	Uneasy in crowds	Restricted Travel Away From Home	Behavior 11
5	DEP_5	Loss of sexual interest	Sexual Dysfunction	Neurovegetative 19
15	DEP_15	Suicidal thoughts	Suicidal Ideation/Suicide Attempt	Mood 33

APPENDIX D1: Cluster Analyses of Gulf War (GW) Dataset with Complete Results for k=2, k=3, and k-4 Solutions, and Higher Level Mapping Symptom Profiles for k=5 Solution

This appendix describes in detail the complete findings of the cluster analyses using the GW dataset for the k=2, k=3, and k=4 solutions. The main text describes the findings in the k=5 solution. Also in this appendix are symptom profiles for all of the four cluster solutions with the symptoms renamed as higher level symptoms (behavioral, interpersonal, mood, neurovegetative, psychotic, and somatic). This was undertaken in an attempt to delineate whether the higher level categories elucidated a pattern for the symptoms that were identified as higher or lower than expected in the symptom profiles. Unfortunately, these analyses did add value to the overall interpretation of theses analyses.

The k=2 solution partitioned the dataset into two groups, both of which contained individuals with diagnoses of PTSD, MDD, and alcohol dependence. Notably, Cluster 2.1 had a much higher mean number of psychiatric diagnoses (mean of 2.0) than Cluster 2.2 (mean of 0.9). Upon reviewing the symptom profiles, Cluster 2.1 was higher across all clinical inventories and expert-derived symptom groupings. These symptom profiles are shown in Figures D1.1a–D1.1i. Statistical significance between the same symptoms in the two different symptom profiles was not expected, given the fact that 161 symptoms were used in the cluster analyses; however, viewing the results in the form of symptom profiles is useful, as these graphs quickly convey the overall stratifications of the groups. The vast majority of the mean values across the symptom profile for Cluster 2.1 were greater than for Cluster 2.2. The only item that overlapped between the two profiles was item 19b of the Beck Depression Inventory, which asked if an individual

was “trying to diet.” The individuals in Cluster 2.1 had a significantly greater proportion of alcohol dependence ($\chi 2=9.3$, $p<.001$) than individuals in Cluster 2.2 (Table 5.1).

Scores in several neuropsychiatric domains were also significantly higher (i.e., better) in Cluster 2.2 than in Cluster 2.1, including measures contained within the Executive Functioning ($p<.05$, $p<.01$), Memory ($p<.05$, $p<.01$, $p<.001$), Performance IQ ($p<.05$, $p<.01$), and Verbal IQ ($p<.05$, $p<.01$) domains (Table 5.2, Figure D1.1j). There was only one significant difference in imaging markers: Specifically, Cluster 2.1 had greater right lenticular white matter than Cluster 2.2 ($p<0.05$; see Table 5.2 and Figure D1.1k). There were no significant differences in the cortisol measures between Clusters 2.1 and 2.2.

Table D1.1. Biomarker Differences Across GW Clusters in the GW Dataset for k=2

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	ANCOVA P-value	Sig	F	Post-Hoc Cluster Difference[TK]
Neuro-psychiatric (D.F.=1,228)								
Executive Functioning 1	9.55E+00	3.02E+00	1.09E+01	3.19E+00	5.43E-03	**	7.88	2>1**
Executive Functioning 2	1.08E+01	2.58E+00	1.19E+01	2.85E+00	9.52E-03	**	6084	2>1*
Memory 1	9.91E+00	3.33E+00	1.13E+01	3.28E+00	6.39E-03	**	305	2>1**
Memory 2	1.10E+01	2.70E+00	1.21E+01	3.03E+00	1.28E-02	*	6.29	2>1*
Memory 4	2.26E+01	7.42E+00	2.54E+01	7.95E+00	1.61E-02	*	5.88	2>1*
Performance IQ 1	6.24E+01	1.41E+01	6.98E+01	1.48E+01	4.25E-04	***	12.79	2>1***
Performance IQ 2	3.83E+01	1.24E+01	4.35E+01	1.18E+01	2.42E-03	**	9.42	2>1**
Verbal IQ 1	4.34E+01	1.09E+01	4.63E+01	9.45E+00	4.00E-02	*	4.27	2>1*
Verbal IQ 3	1.81E+01	4.13E+00	1.99E+01	4.19E+00	2.65E-03	**	9.27	2>1**
Imaging Volumes (D.F.=1,165)								
Right Lenticular White Matter	2.44E-03	6.00E-04	2.22E-03	4.00E-04	9.00E-03	**	6.8	1>2*
Right Frontal CSF	4.26E-02	9.31E-03	4.52E-02	8.14E-03	4.20E-02	*	4.19	N.S.
Right Parietal CSF	2.13E-02	4.98E-03	2.29E-02	4.97E-03	4.28E-02	*	4.17	N.S.
Right Temporal CSF	1.42E-02	2.68E-03	1.52E-02	2.61E-03	2.27E-02	*	5.29	N.S.
Left Temporal CSF	1.29E-02	2.38E-03	1.39E-02	3.16E-03	2.63E-02	*	5.03	N.S.

D.F.: degrees of freedom
S.D.: standard deviation
N.S.: not significant
TK: Tukey-Kramer
*, p<.05; **, p<.01; ***, p<.001

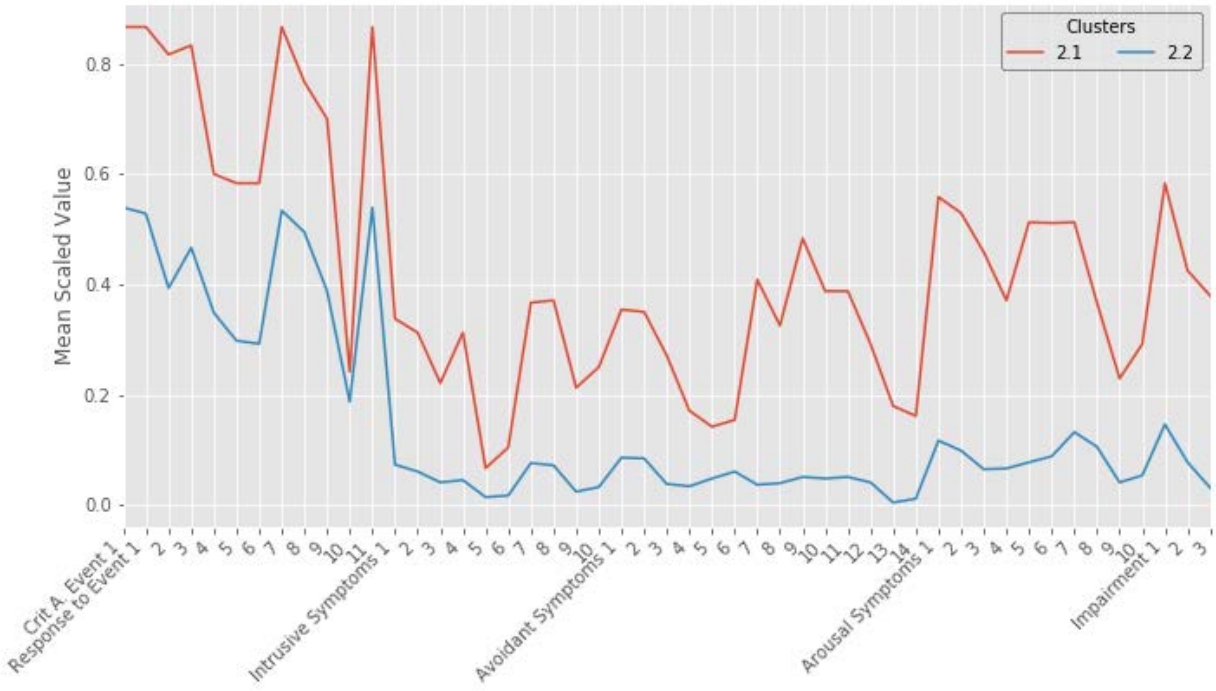

Figure D1.1a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=2)

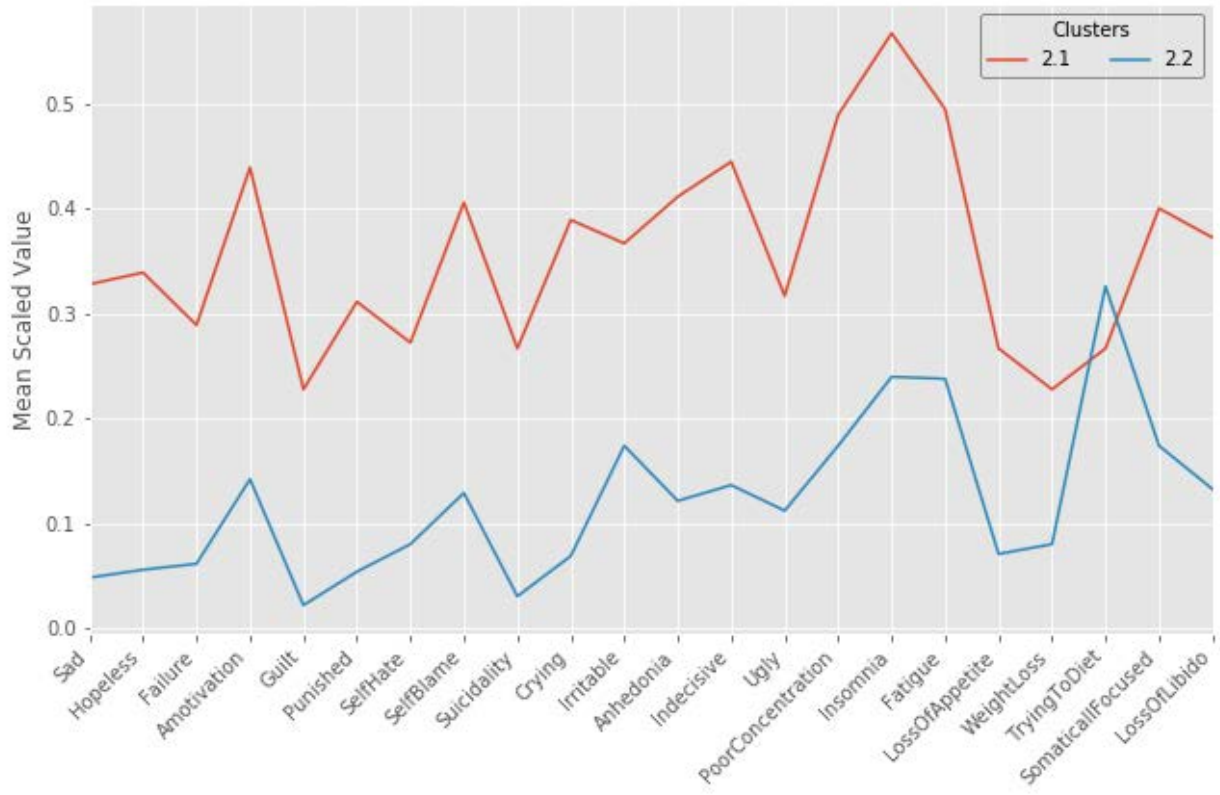

Figure D1.1b. Symptom Profiles for All Items in the Beck Depression Inventory (k=2)

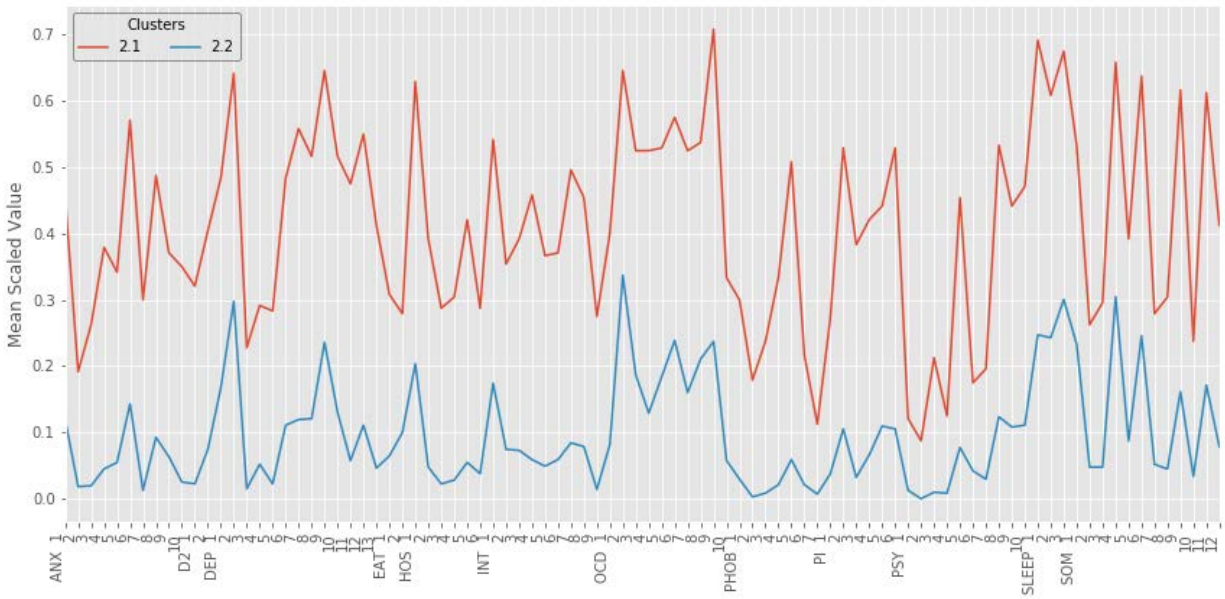

Figure D.1.1c. Symptom Profiles for All Items in the Symptom Checklist-90 (k=2)

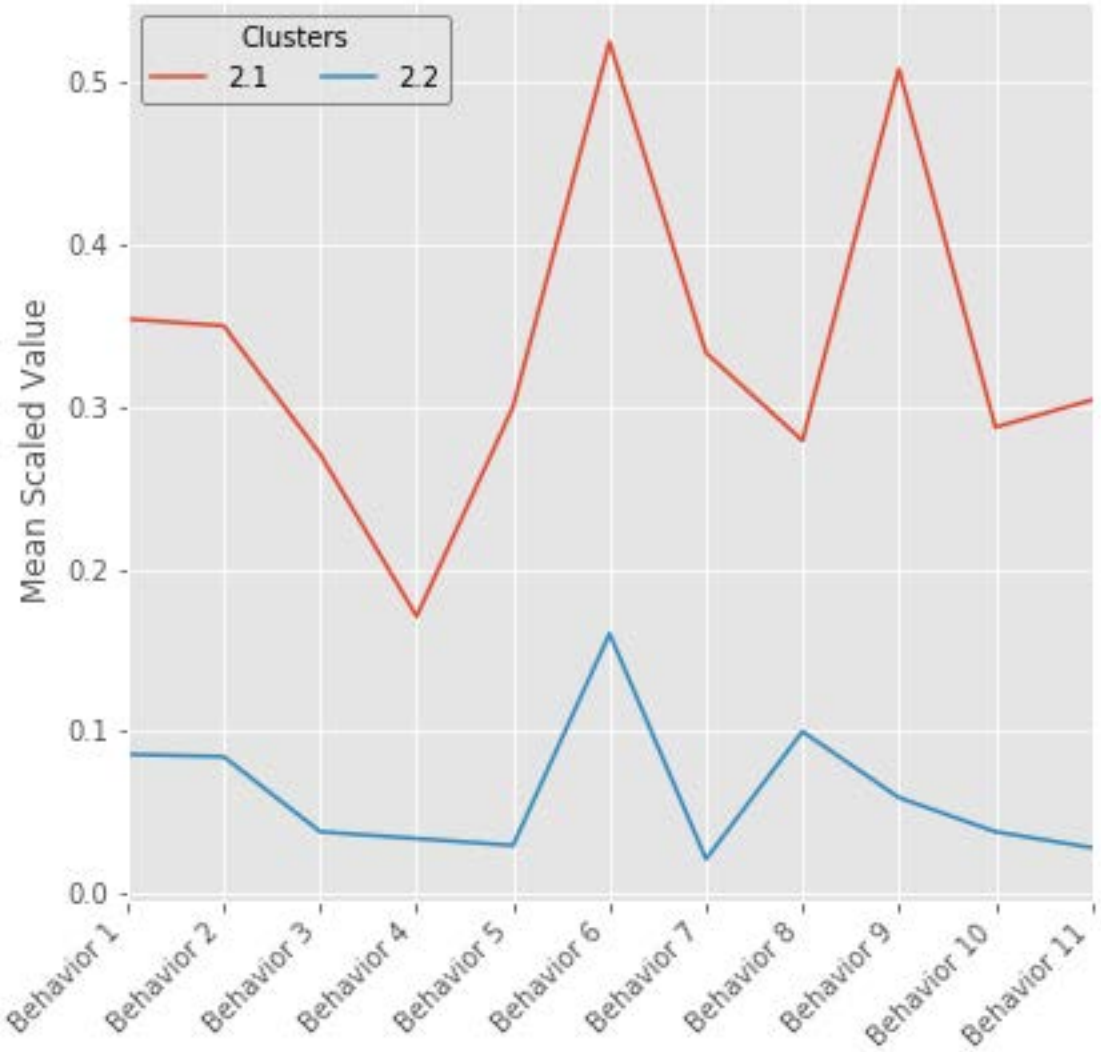

Figure D1.1d. Symptom Profiles for All Behavioral Items (k=2)

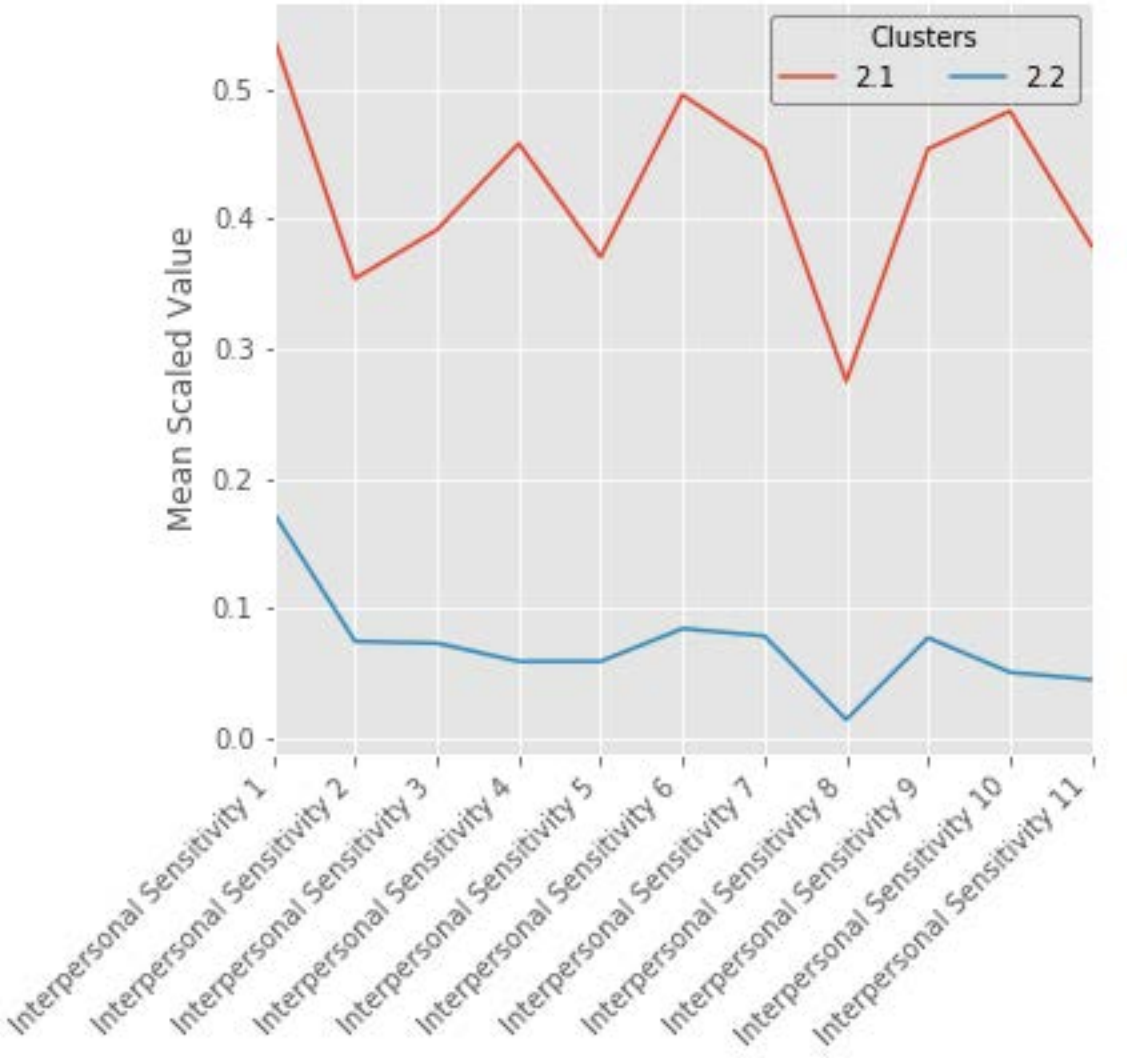

Figure D1.1e. Symptom Profiles for All Interpersonal Sensitivity Items (k=2)

Figure D1.1f. Symptom Profiles for All Mood Items (k=2)

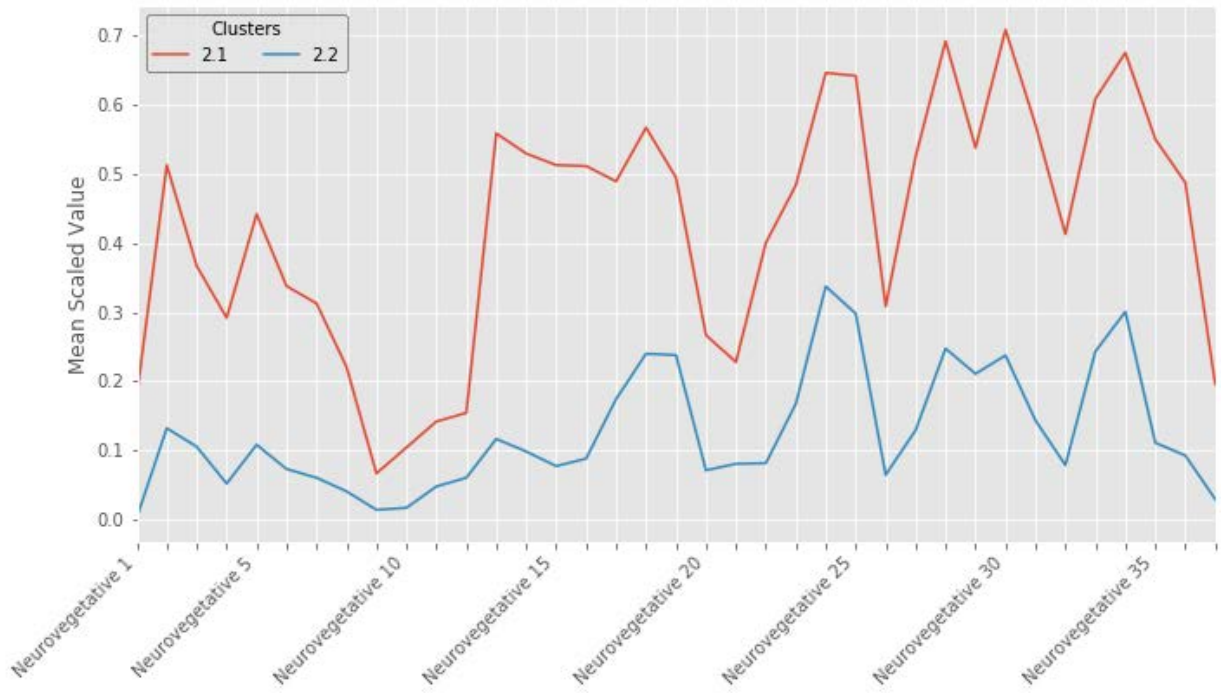

Figure D1.1g. Symptom Profiles for All Neurovegetative Items (k=2)

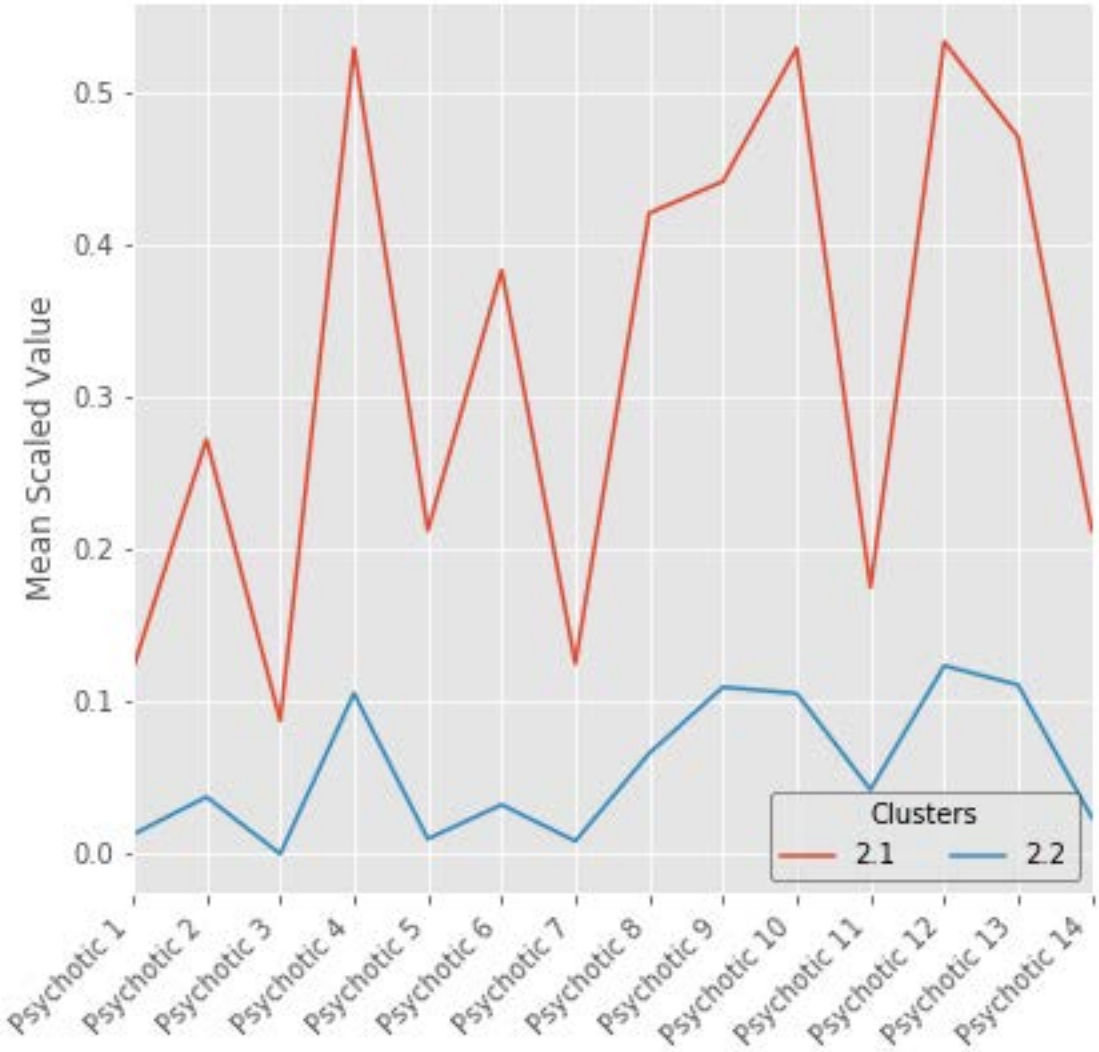

Figure D1.1h. Symptom Profiles for All Psychotic Items (k=2)

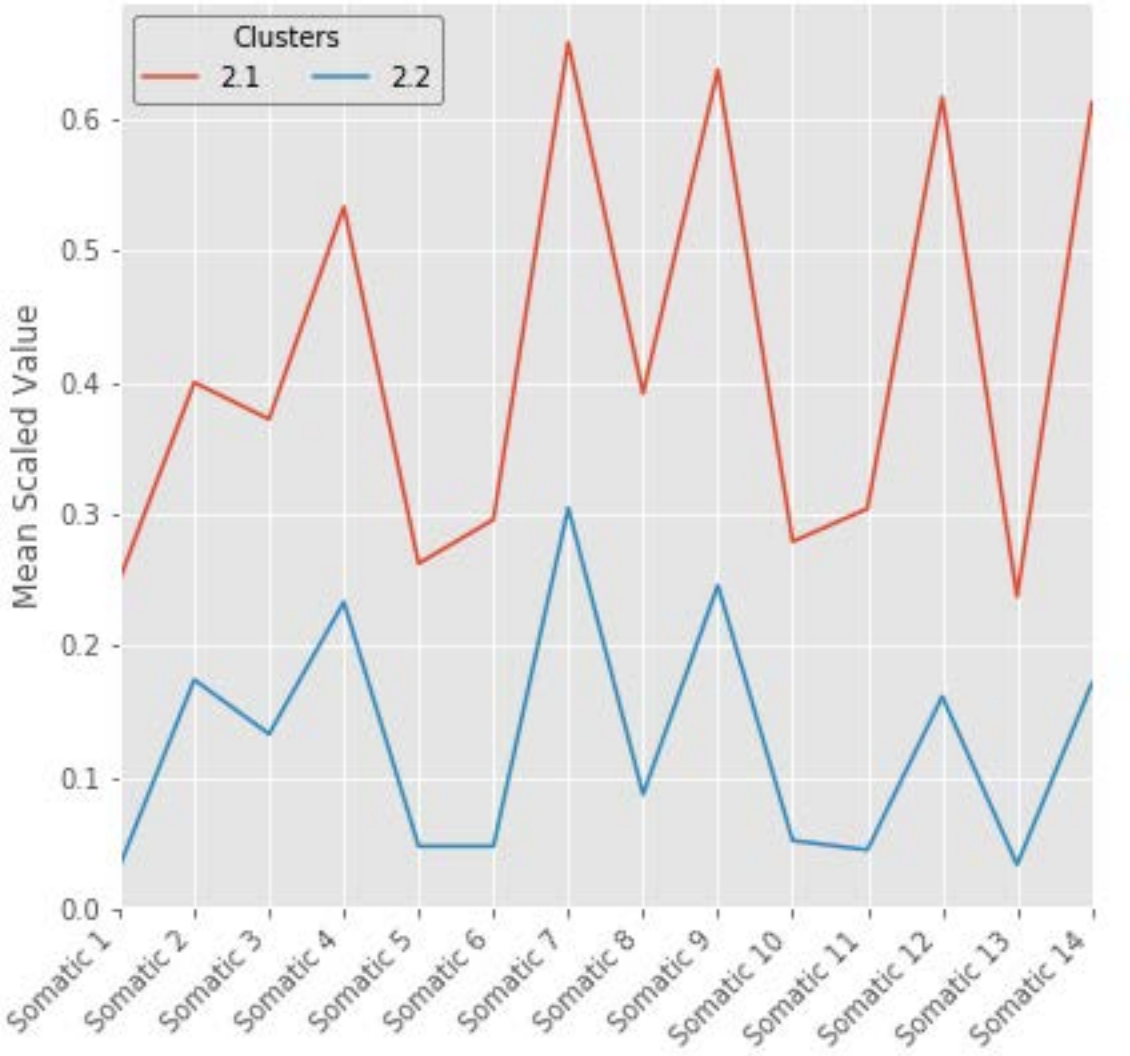

Figure D1.1i. Symptom Profiles for All Somatic Items (k=2)

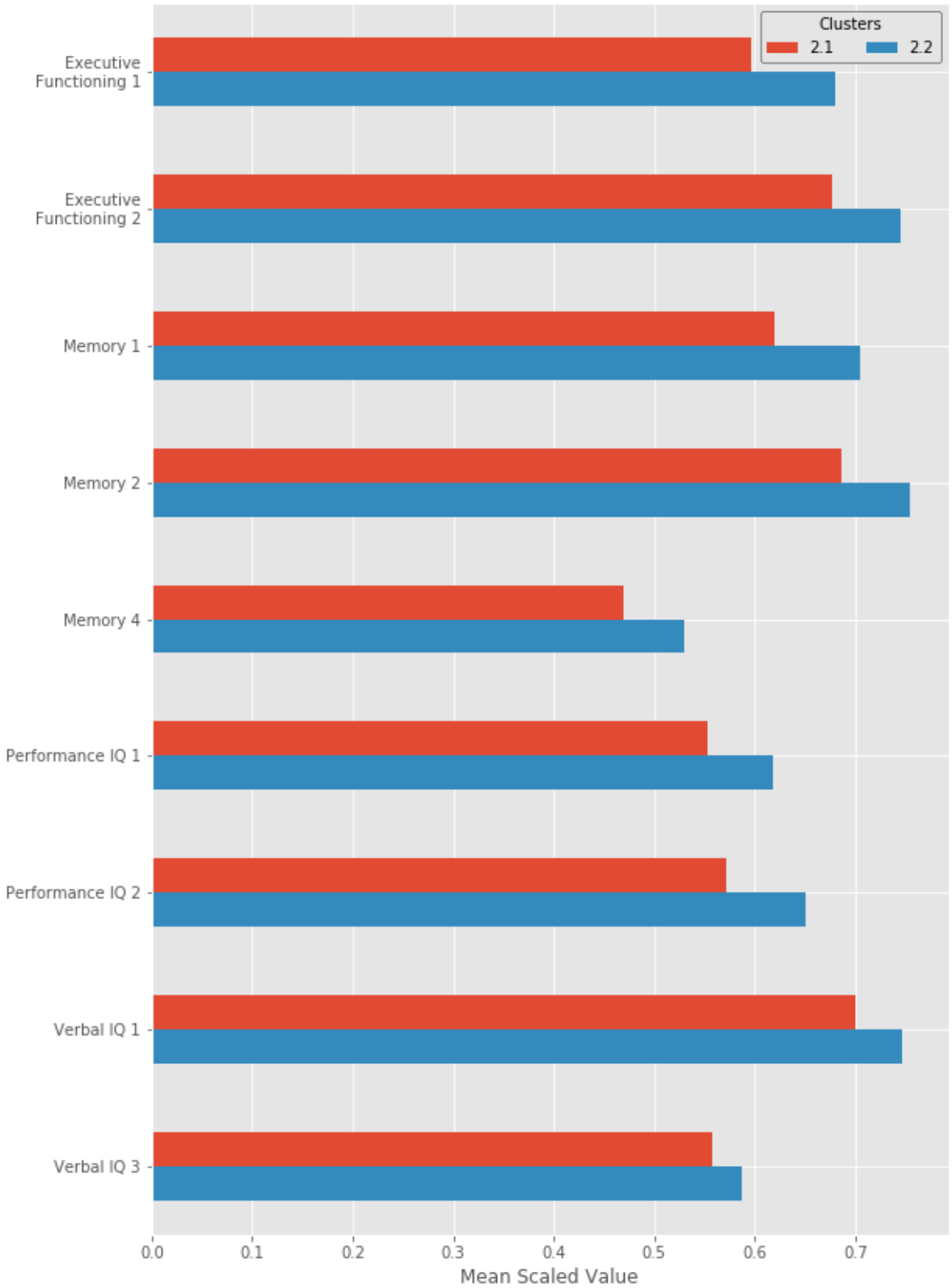

Figure D1.1j. Neuropsychiatric Markers with Significant Differences Across Clusters (k=2)

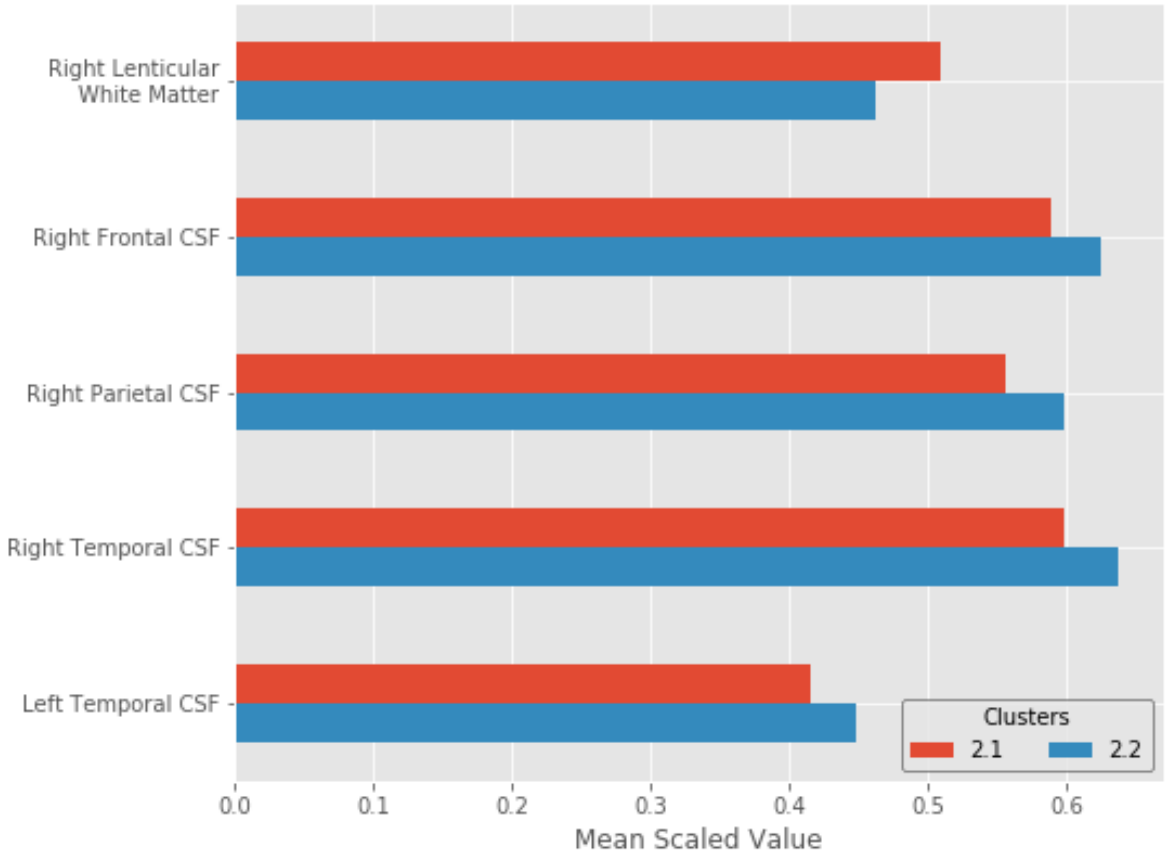

Figure D1.1k. Imaging Markers with Significant Differences Across Clusters (k=2)

The k=3 solution was not identified as an optimal clustering solution by either CH or wss criteria, and had a midlevel ARI of 0.56. This solution did produce stable clusters, however, once bootstrapping was applied (>96 percent stable clusters). Within this solution, Cluster 3.1 comprised 14 individuals, 12 with PTSD, seven with MDD, and five with alcohol dependence (Table D1.1). This cluster had the highest average number of DSM diagnoses (DSM avg=2.9 vs. DSM avg=2.4 for Cluster 3.2 and DSM avg=1.2 for Cluster 3.3). However, 20 of 70 individuals in Cluster 3.2 had PTSD, 13 had MDD, and 28 had alcohol dependence. Consistently, individuals in Cluster 3.1 reported the highest level of symptom distress across all items in both clinical scales, as well as the expert-identified symptom groupings (Table D1.2 and Figures D1.2a–D1.2i). The incidence of alcohol dependence across all three clusters differed significantly ($\chi 2=13.7$, $p<.001$; Table 5.1), with the percentage of individuals with alcohol dependence being the greatest in 3.1 and least in 3.3, decreasing as the symptom profiles lowered.

Individuals in Cluster 3.1 also performed significantly worse on measures of Executive Functioning ($p<.05$), Memory ($p<.05$), and Verbal IQ than individuals in Cluster 3.3. Individuals in Cluster 3.1 performed significantly worse on Verbal IQ (Verbal IQ 3) than individuals in either Cluster 3.2 ($p<0.01$) or Cluster 3.3 ($p<0.001$, Figure D1.2h). Individuals in Cluster 3.2 also performed worse than Cluster 3.3 in the domain of Performance IQ ($p<.05$).

Within the imaging data, Cluster 3.1 had significantly larger average left parietal cortex volume and smaller average left and right parietal CSF volumes than Cluster 3.3 ($p<0.01$ for all comparisons) and right lenticular white matter ($p<0.01$ for all comparisons). Furthermore, Cluster 3.1 had a larger average right lenticular white matter volume than Cluster 3.2 ($p<0.05$ and Cluster 3.3 ($p<.01$)

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	Cluster 3 Mean	S.D.	ANCOVA P-value	Sig	F	Post Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=2,227)										
Executive Functioning 1	8.23E+00	2.86E+00	1.03E+01	3.43E+00	1.08E+01	-3.03E+00	1.09E-02	*	4.61	3>1*
Executive Functioning 2	9.92E+00	2.22E+00	1.13E+01	3.04E+00	1.19E+01	2.71E+00	2.04E-02	*	3.96	3>1*
Memory 1	8.92E+00	2.36E+00	1.05E+01	3.85E+00	1.13E+01	3.08E+00	1.63E-02	*	4.19	3>1*
Memory 2	1.00E+01	2.31E+00	1.14E+01	3.40E+00	1.21E+01	2.77E+00	1.90E-02	*	4.04	3>1*
Performance IQ 1	6.06E+01	1.39E+01	6.42E+01	1.37E+01	7.03E+01	1.50E+01	1.63E-03	**	6.06	3>2*
Performance IQ 2	3.48E+01	1.10E+01	4.00E+01	1.28E+01	4.39E+01	1.16E+01	3.19E-03	**	5.89	3>1*
Verbal IQ 1	3.81E+01	1.16E+01	4.47E+01	1.04E+01	4.66E+01	9.25E+00	4.85E-03	**	5.45	3>1**
Verbal IQ 3	1.52E+01	4.18E+00	1.91E+01	3.78E+00	1.99E+01	4.25E+00	1.01E-04	***	9.58	3>1***,2>1**
Imaging Volumes (D.F.=2,164)										
Left Parietal Cortex	4.81E-02	-4.50E-03	4.53E-02	-3.70E-03	4.50E-02	-3.60E-03	1.22E-02	*	4.53	1>3*
Right Lenticular White Matter	2.71E-03	8.19E-04	2.25E-03	4.45E-04	2.24E-03	4.51E-04	1.12E-02	**	4.62	1>2*, 1>3**
Right Frontal CSF	4.52E-02	8.23E-03	4.42E-02	9.13E-03	3.92E-02	6.84E-03	4.44E-02	*	3.17	N.S.
Left Frontal CSF	4.48E-02	7.71E-03	4.46E-02	8.25E-03	3.91E-02	6.76E-03	3.46E-02	*	3.43	N.S.
Right Parietal CSF	1.91E-02	-4.30E-03	2.20E-02	-4.80E-03	2.30E-02	5.10E-03	2.52E-02	*	3.77	3>1*
Left Parietal CSF	1.83E-02	-4.30E-03	2.18E-02	-4.90E-03	2.21E-02	-4.90E-03	2.77E-02	*	3.67	3>1*
Left Occipital CSF	5.22E-03	1.41E-03	4.96E-03	1.18E-03	4.27E-03	1.24E-03	4.77E-02	*	3.1	N.S.

D.F., degrees of freedom
S.D., standard deviation
N.S., not significant
TK, Tukey-Kramer
*, p<.05; **, p<.01; ***, p<.001

Table D1.2. Biomarker Differences Across Clusters in the GWA Dataset for k=3

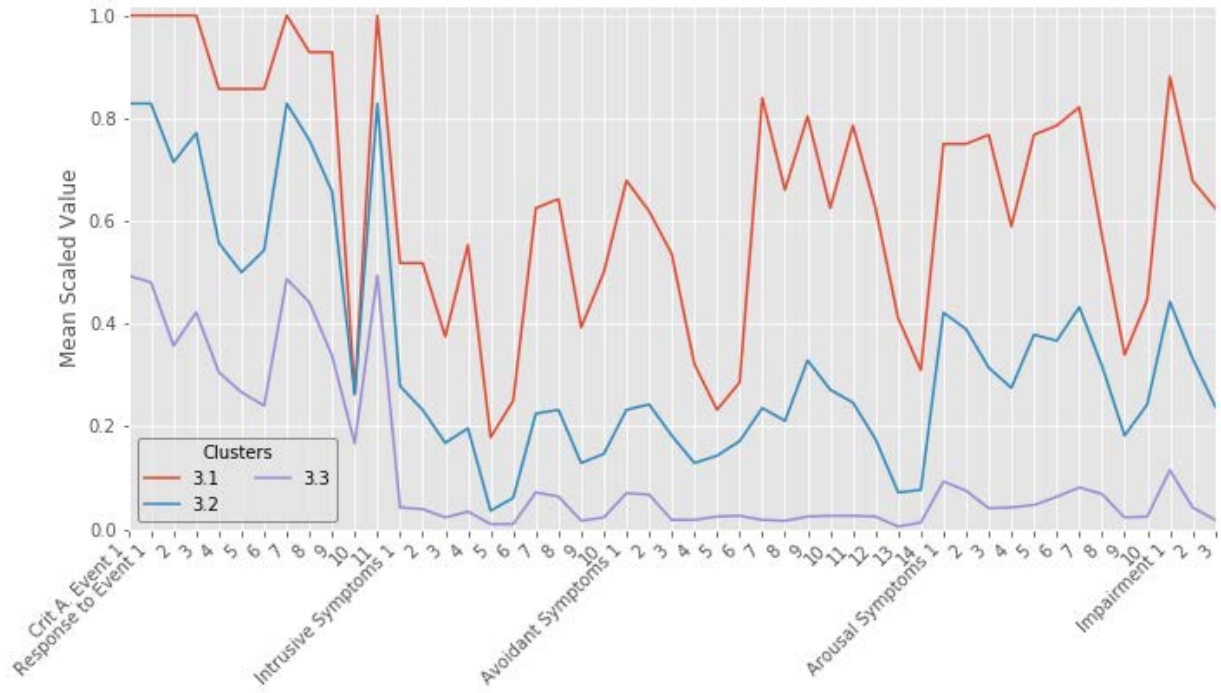

Figure D1.2a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=3)

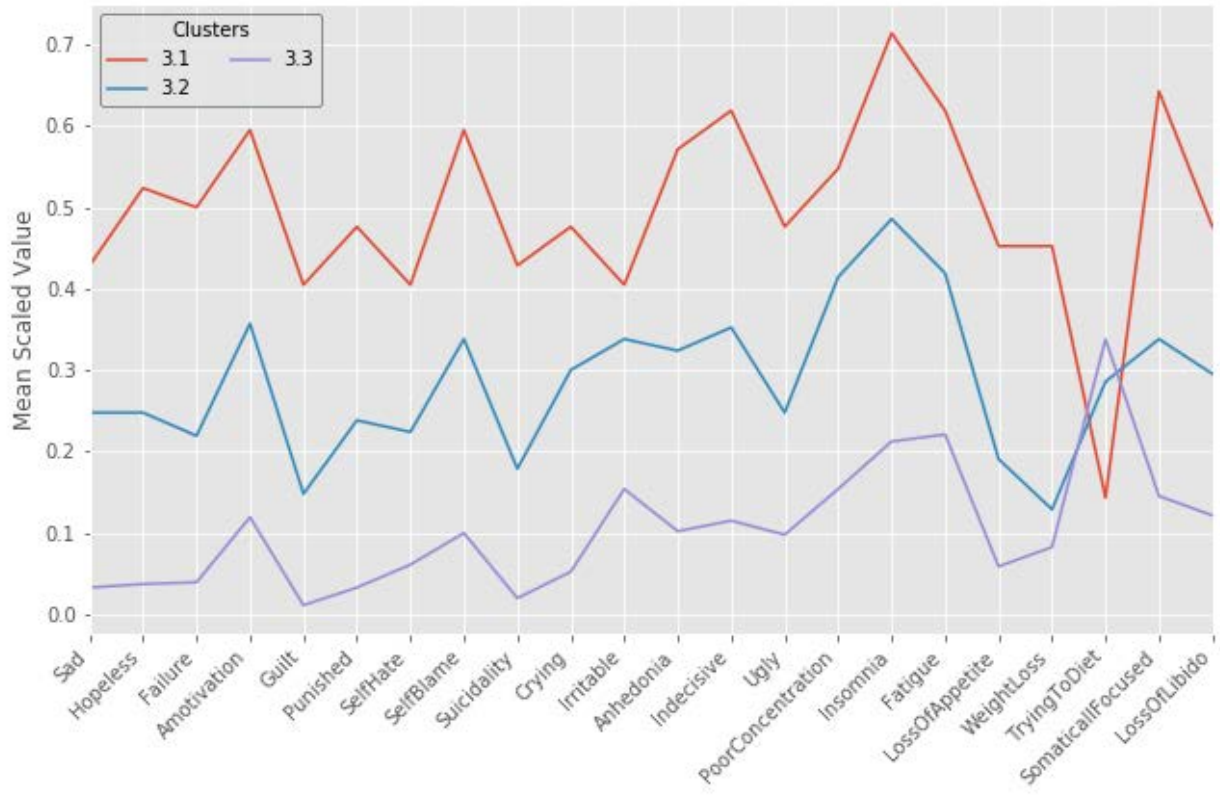

Figure D1.2b. Symptom Profiles for All Items in the Beck Depression Inventory (k=3)

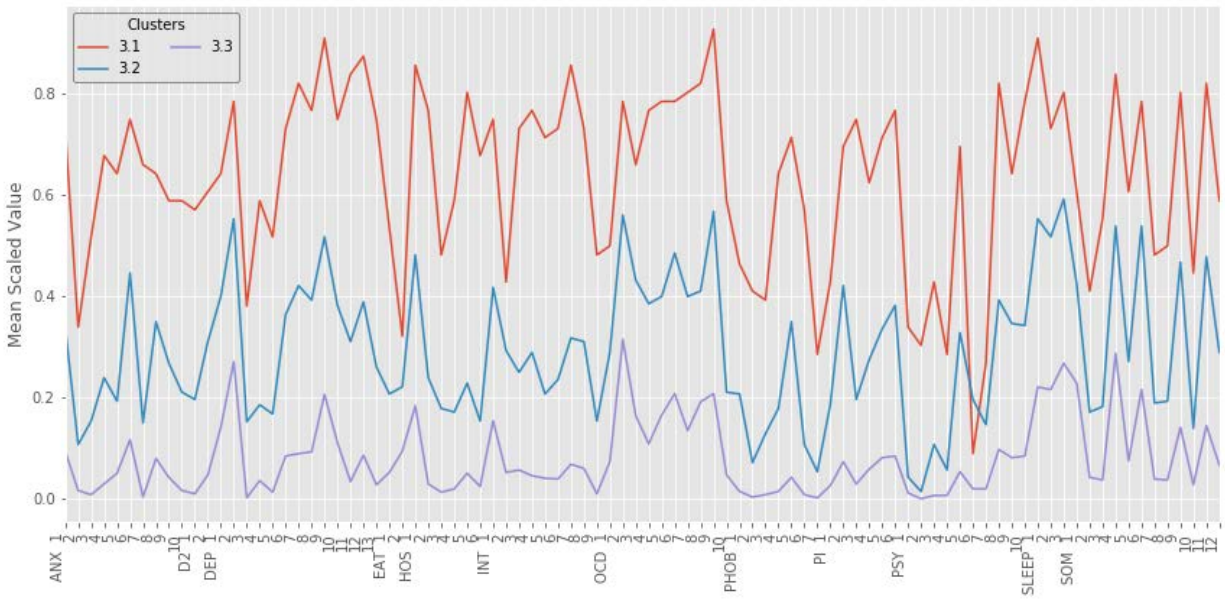

Figure D1.2c. Symptom Profiles for All Items in the Symptom Checklist 90 (k=3)

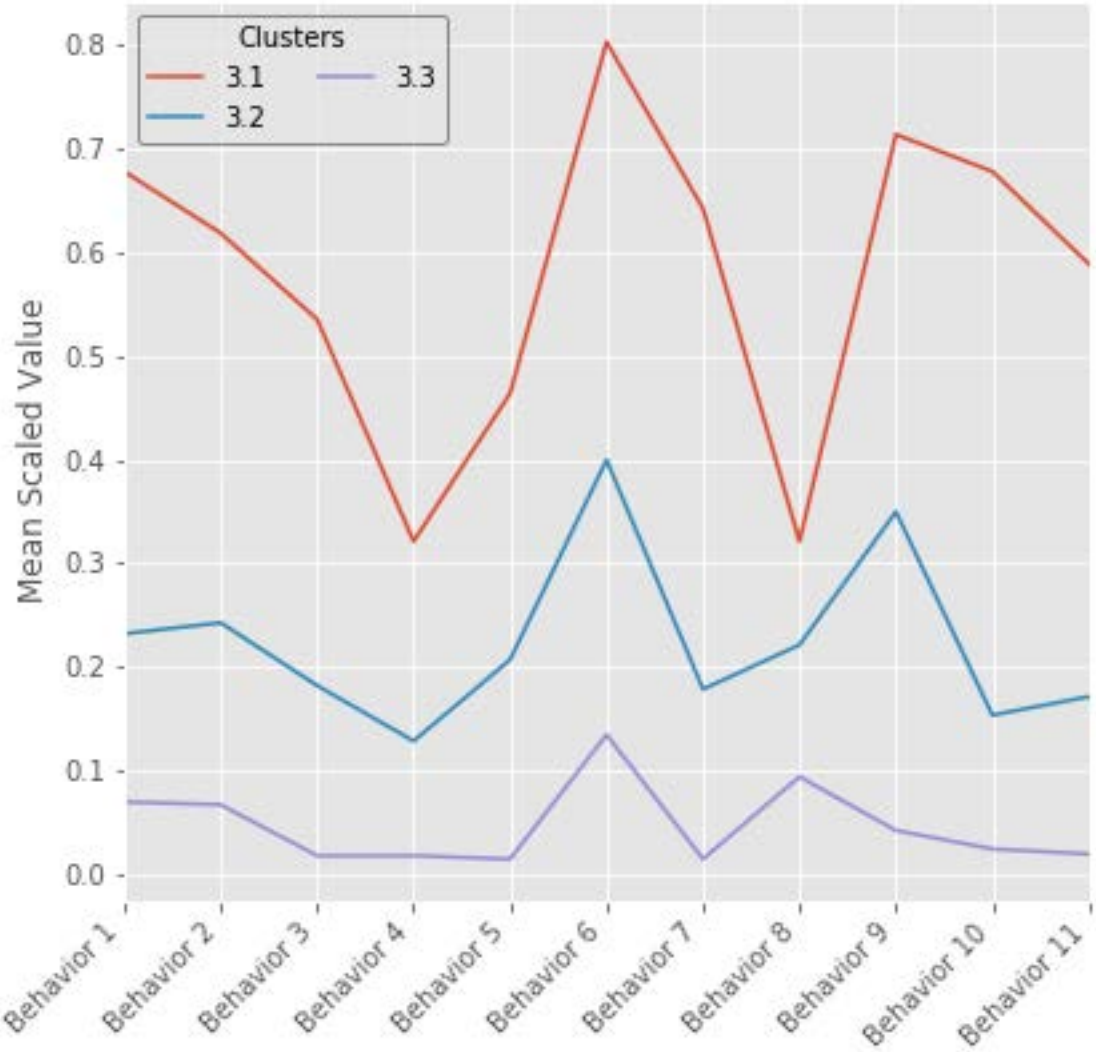

Figure D1.2d. Symptom Profiles for All Behavioral Items (k=3)

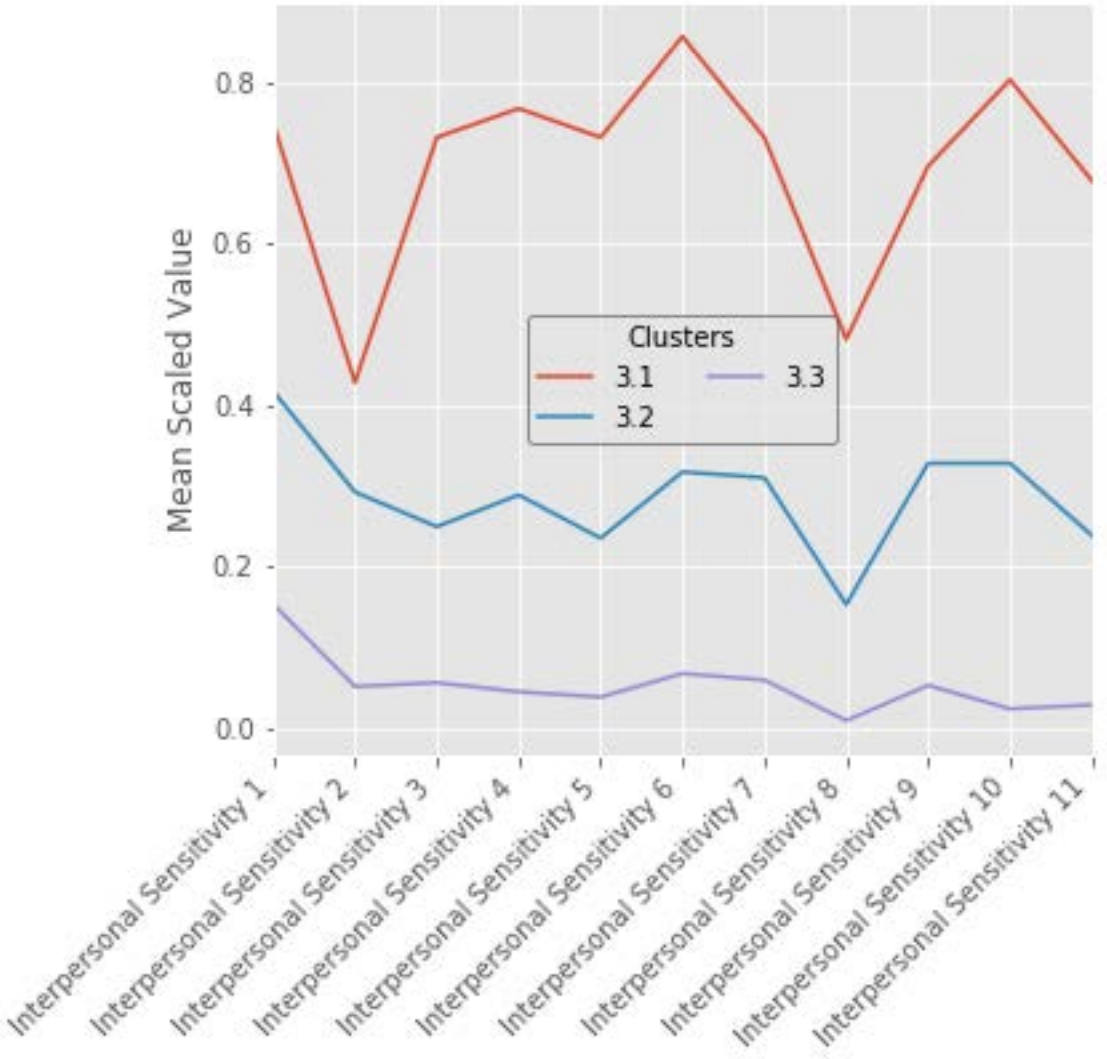

Figure D1.2e. Symptom Profiles for All Interpersonal Sensitivity Items (k=3)

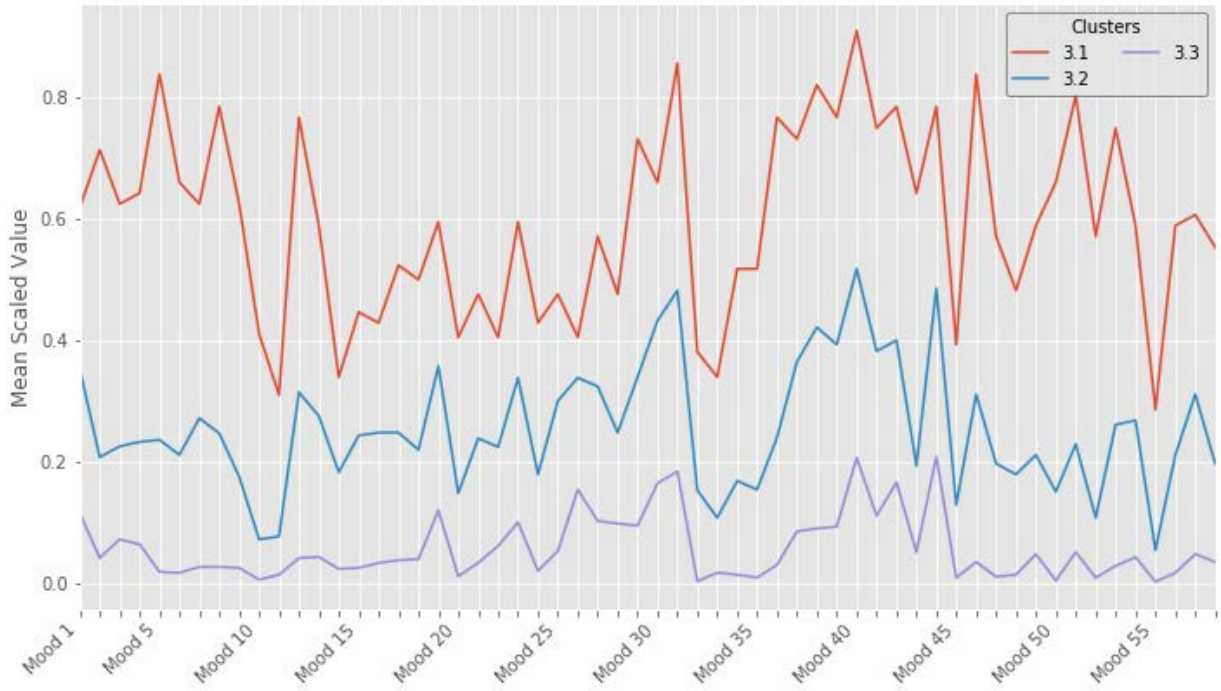

Figure D1.2f. Symptom Profiles for All Mood Items (k=3)

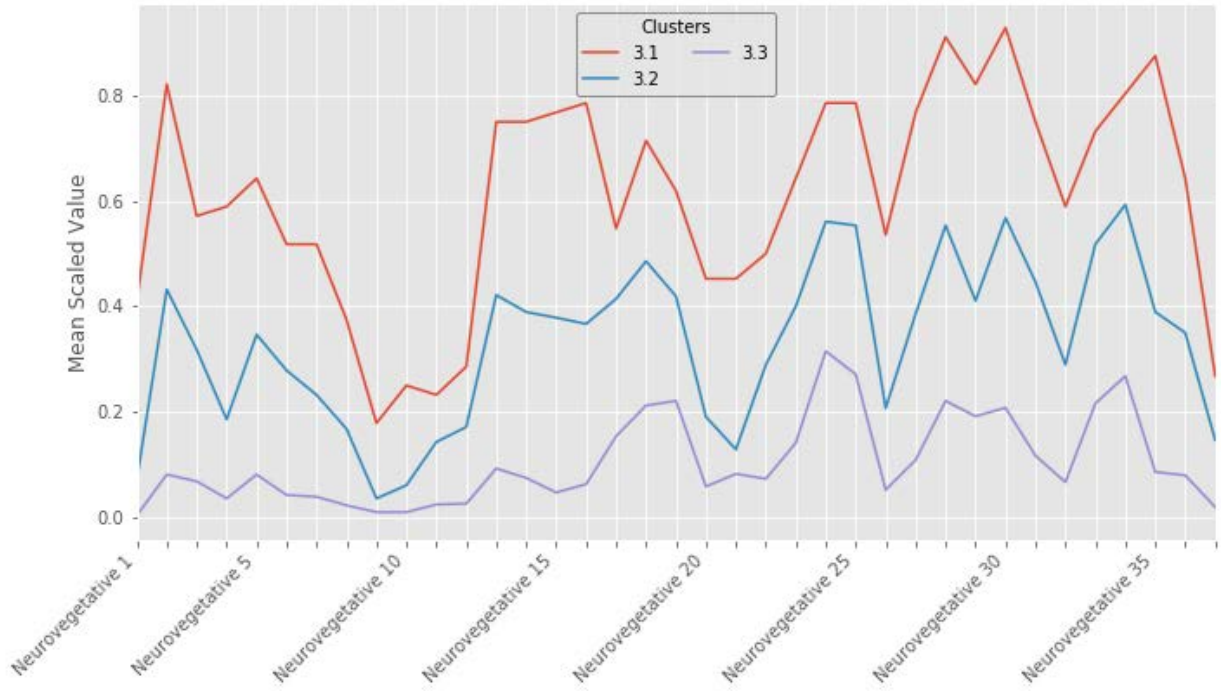

Figure D1.2g. Symptom Profiles for All Neurovegetative Items (k=3)

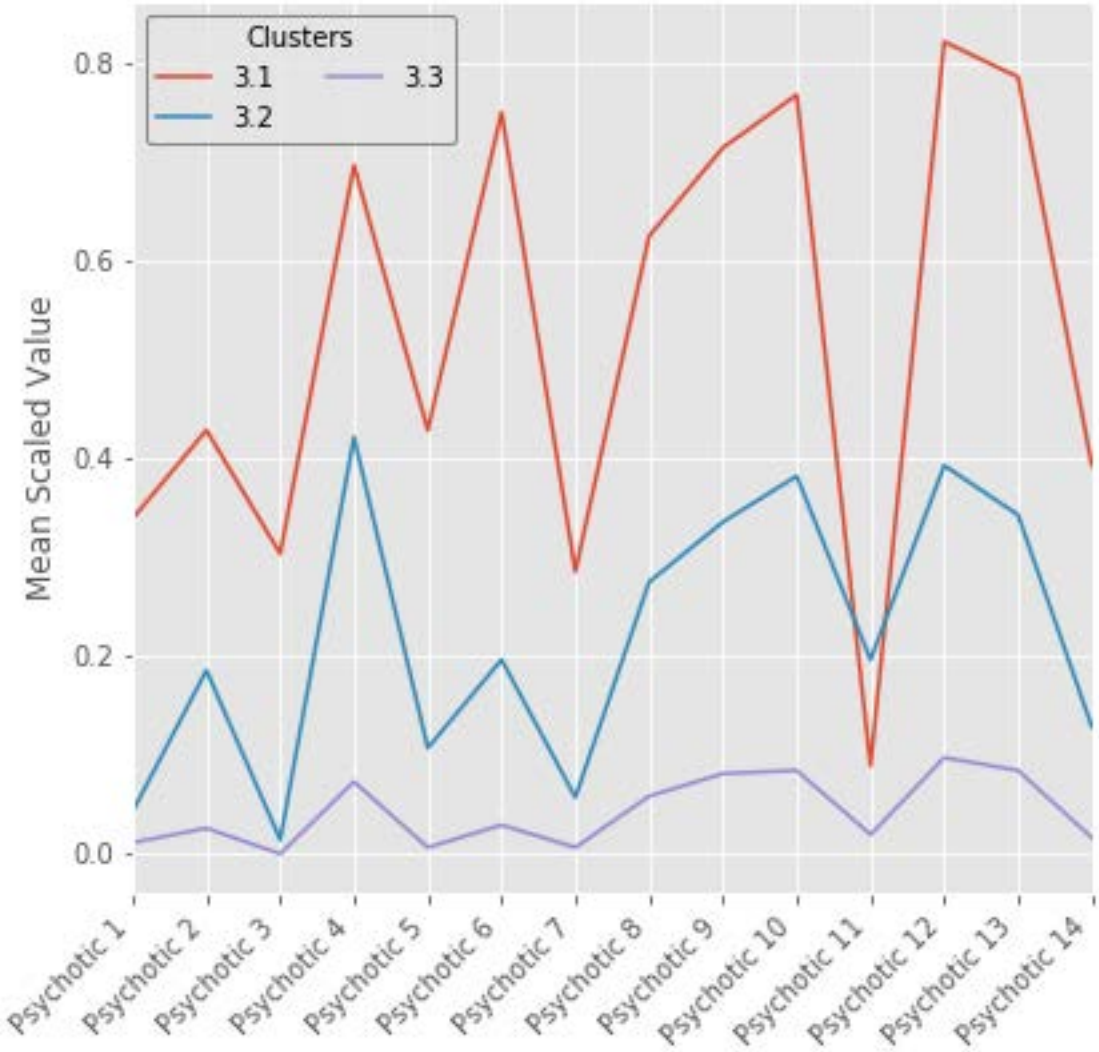

Figure D1.2h. Symptom Profiles for All Psychotic Items (k=3)

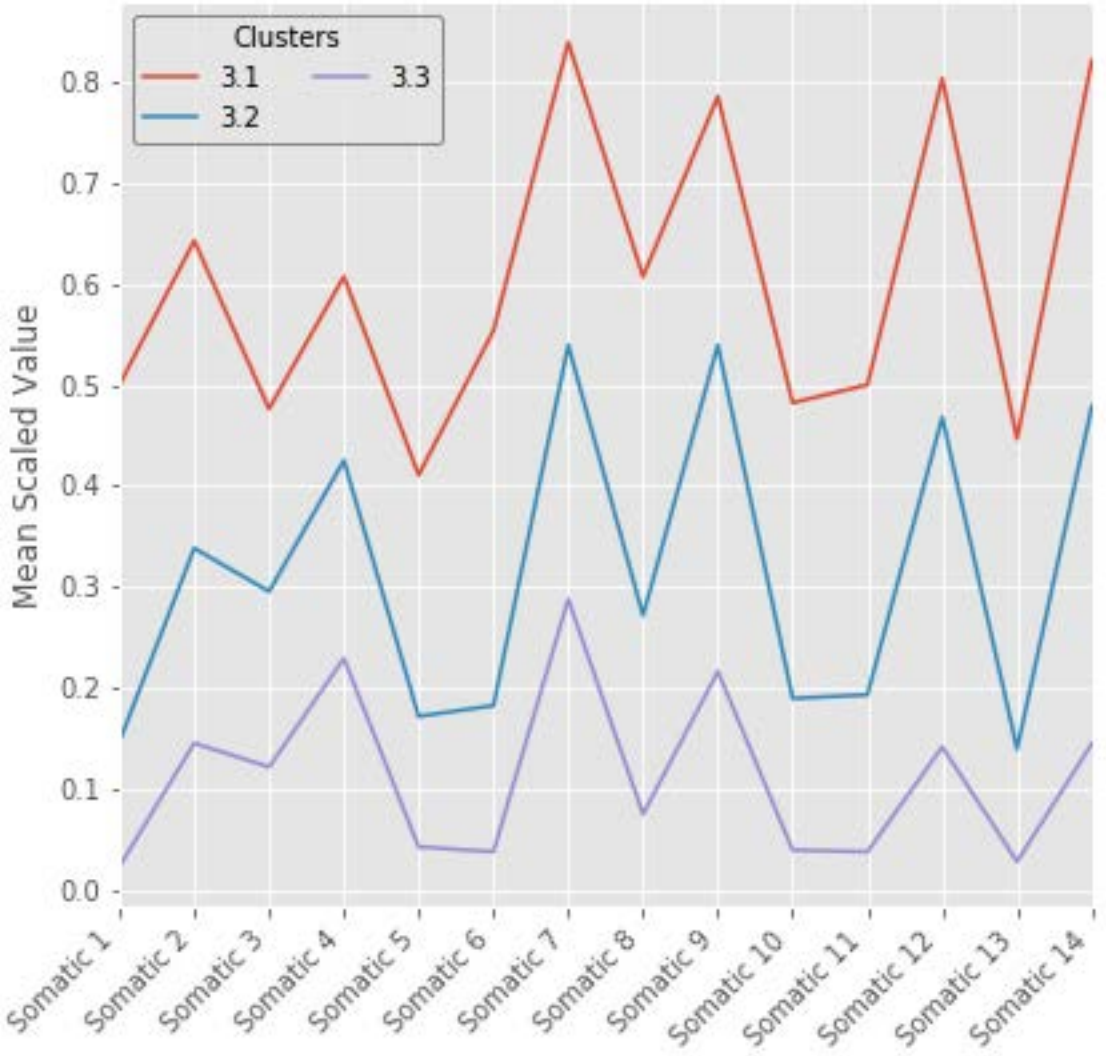

Figure D1.2i. Symptom Profiles for All Somatic Items (k=3)

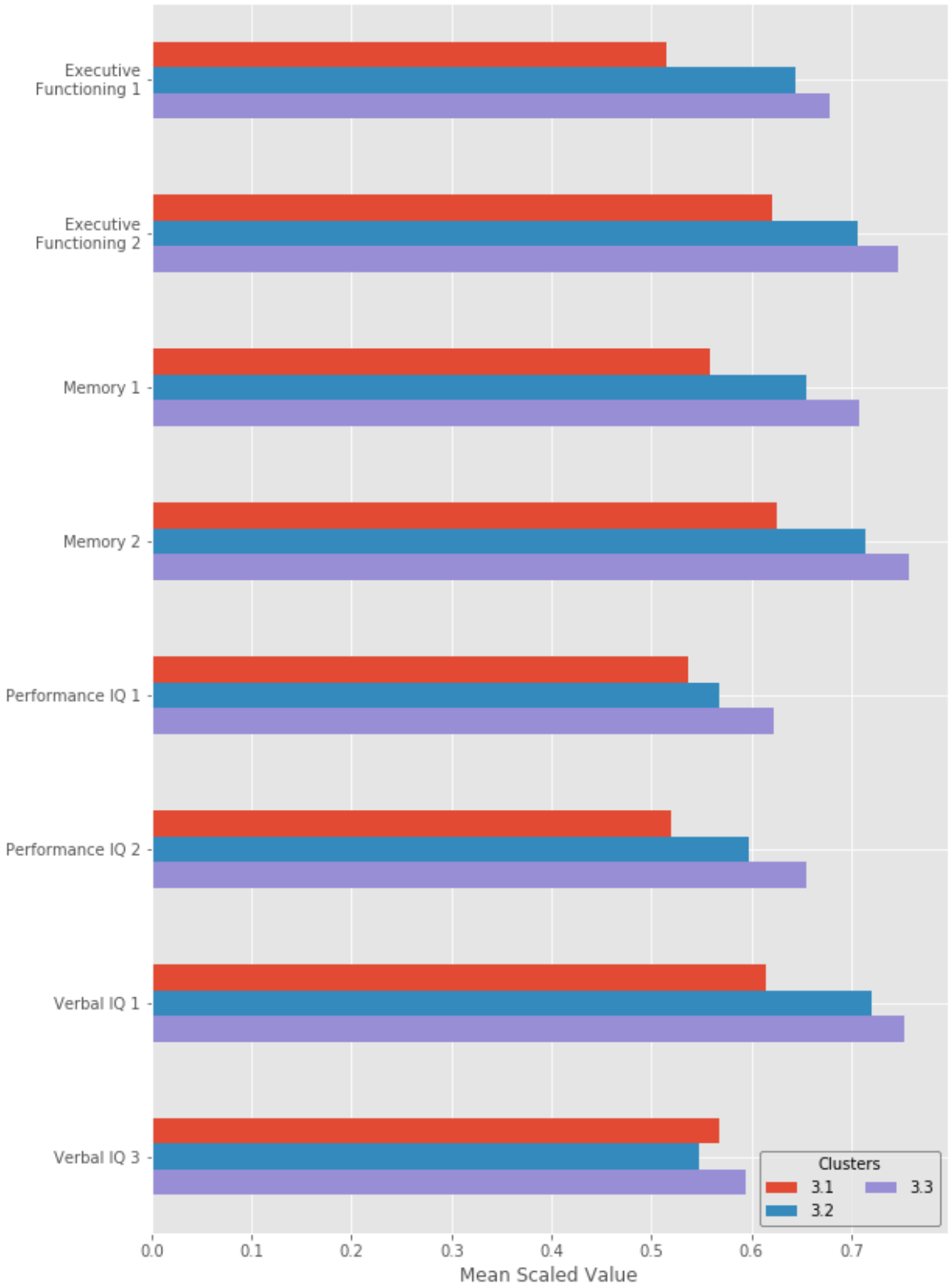

Figure D1.2j. Neuropsychiatric Markers with Significant Differences Across Clusters (k=3)

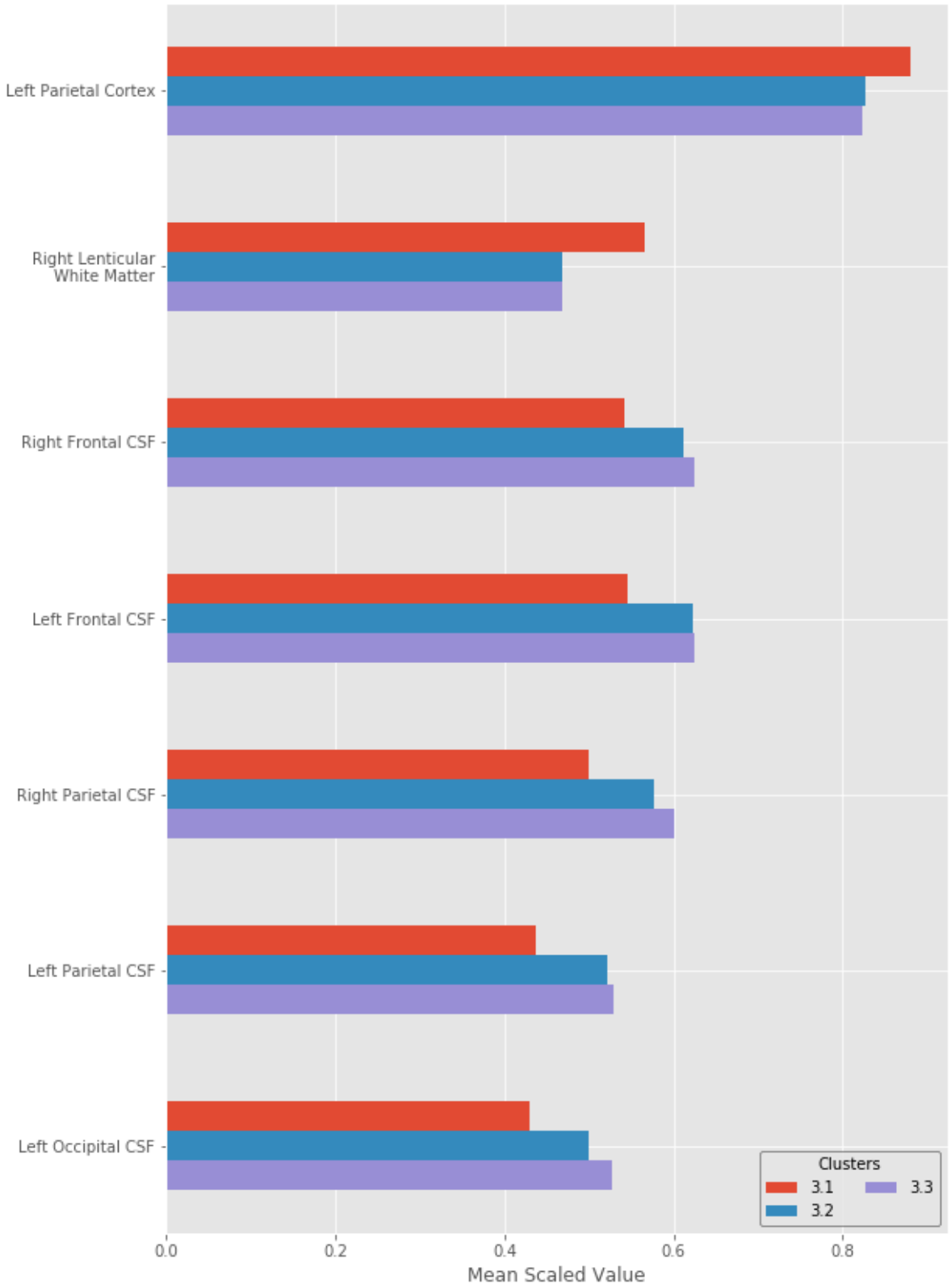

Figure D1.2k. Imaging Markers with Significant Differences Across Clusters (k=3)

Interestingly, the individuals in Cluster 3.1 in the k=3 solution continued to cluster together in the k=4 (Cluster 4.1) and k=5 (Cluster 5.1) cluster solutions. The k=4 cluster solution had a relatively low ARI (0.39) and lower cluster stability, as demonstrated through bootstrapping (Table 5.1). This solution continued to report that Cluster 4.1 (akin to Cluster 3.1) had the highest level of symptom distress across all clusters (Figures D1.4a–i), followed by Cluster 4.2, which included most of the remaining individuals with a PTSD diagnosis (15/33). The incidence of alcohol dependence across all four clusters differed significantly ($\chi 2=20.3$, $p<.0001$; Table 5.1), with the percentage of individuals with alcohol dependence being the greatest in 4.1, and least in 4.4, generally decreasing as the symptom profiles lowered. This decreasing pattern as the cluster symptom profile decreased was also seen in the incidence of childhood trauma ($\chi 2=14.4$, $p<.001$; Table 5.1).

Individuals in Cluster 4.1 performed significantly worse, on average, than those in the other clusters on tests of the following neuropsychiatric domains: Executive Functioning ($p<.05$), Performance IQ ($p<.05$, $p<.01$), and Verbal IQ ($p<.05$, $p<.01$, $p<.001$) (Table D1.3 Figure D1.3j). Within the imaging markers, Cluster 4.1 had a larger average left parietal cortex than Cluster 4.4, and larger right lenticular white matter volumes than Clusters 4.3 and 4.4 ($p<.05$ for all; Table D1.3 Figure D1.3k). There were no significant differences between groups in any of the cortisol measures.

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	Cluster 3 Mean	S.D.	Cluster 4 Mean	S.D.	ANCOVA P-value	Sig	F	Post Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=3,226)												
Executive Functioning 1	8.23E+00	2.86E+00	1.03E+01	3.10E+00	1.09E+01	3.31E+00	1.07E+01	3.04E+00	3.19E-02	*	2.99	3>1*,4>1*
Performance IQ 1	6.06E+01	1.39E+01	6.37E+01	1.41E+01	6.89E+01	1.62E+01	7.09E+01	1.31E+01	5.68E-03	**	4.29	4>2*
Performance IQ 2	3.48E+01	1.10E+01	3.99E+01	1.30E+01	4.21E+01	1.19E+01	4.52E+01	1.12E+01	4.36E-03	**	4.49	4>1**
Verbal IQ 1	3.81E+01	1.16E+01	4.55E+01	1.02E+01	4.61E+01	8.96E+00	4.62E+01	1.01E+01	3.14E-02	*	3	3>1*,4>1**
Verbal IQ 3	1.52E+01	4.18E+00	1.92E+01	3.80E+00	1.99E+01	4.08E+00	1.98E+01	4.38E+00	5.42E-04	***	6.07	2>1***, 3>1***, 4>1**
Imaging Volumes (D.F.=3,163)												
Left Parietal Cortex	4.81E-02	4.50E-03	4.54E-02	3.50E-03	4.53E-02	3.70E-03	4.47E-02	3.50E-03	2.07E-02	*	3.34	1>4*
Right Lenticular White Matter	2.71E-03	8.19E-04	2.31E-03	4.68E-04	2.22E-03	4.51E-04	2.23E-03	4.33E-04	1.92E-02	*	3.4	1>4*,1>3*
Right Frontal CSF	4.44E-02	7.46E-03	4.41E-02	9.22E-03	3.92E-02	6.84E-03	4.61E-02	9.14E-03	4.70E-02	**	2.71	N.S.
Right Parietal CSF	1.91E-02	4.30E-03	2.22E-02	4.80E-03	2.20E-02	5.00E-03	2.39E-02	5.00E-03	9.96E-03	**	3.91	4>1*
Left Parietal CSF	1.83E-02	4.30E-03	2.17E-02	4.70E-03	2.14E-02	4.50E-03	2.30E-02	5.30E-03	1.17E-02	*	3.78	4>1*
Right Occipital CSF	4.95E-03	1.24E-03	5.33E-03	1.41E-03	4.51E-03	1.30E-03	5.57E-03	1.39E-03	2.18E-02	*	3.3	N.S.
Left Occipital CSF	4.27E-03	1.20E-03	5.03E-03	1.20E-03	4.77E-03	1.20E-03	5.66E-03	1.50E-03	1.66E-04	***	7.08	4>1*, 4>3**
Right Temporal CSF	1.40E-02	2.60E-03	1.46E-02	2.70E-03	1.44E-02	2.10E-03	1.59E-02	2.90E-03	9.94E-04	***	5.69	4>3**
Left Temporal CSF	1.33E-02	3.12E-03	1.33E-02	2.32E-03	1.23E-02	2.32E-03	1.46E-02	3.20E-03	6.77E-03	**	4.2	N.S.

D.F., degrees of freedom
S.D., standard deviation
N.S., not significant
TK, Tukey-Kramer
*, p<.05; **, p<.01; ***, p<.001

Table D1.3. Biomarker Differences Across Clusters in GWA Dataset for k=4

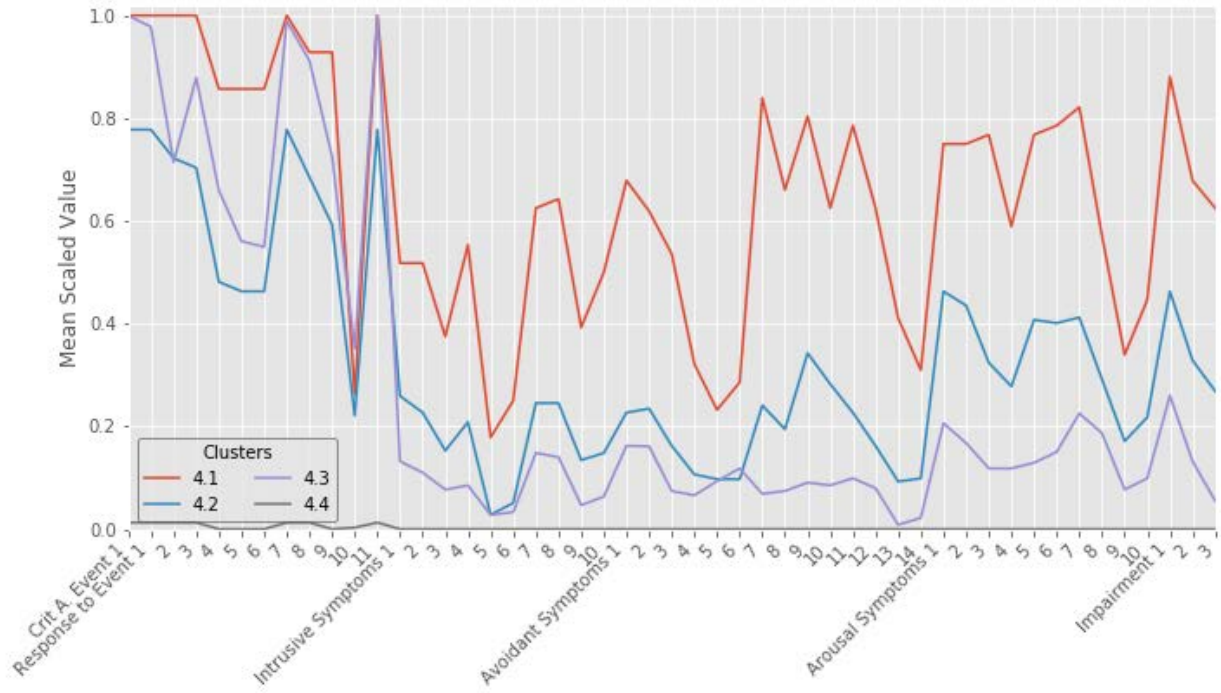

Figure D1.3a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=4)

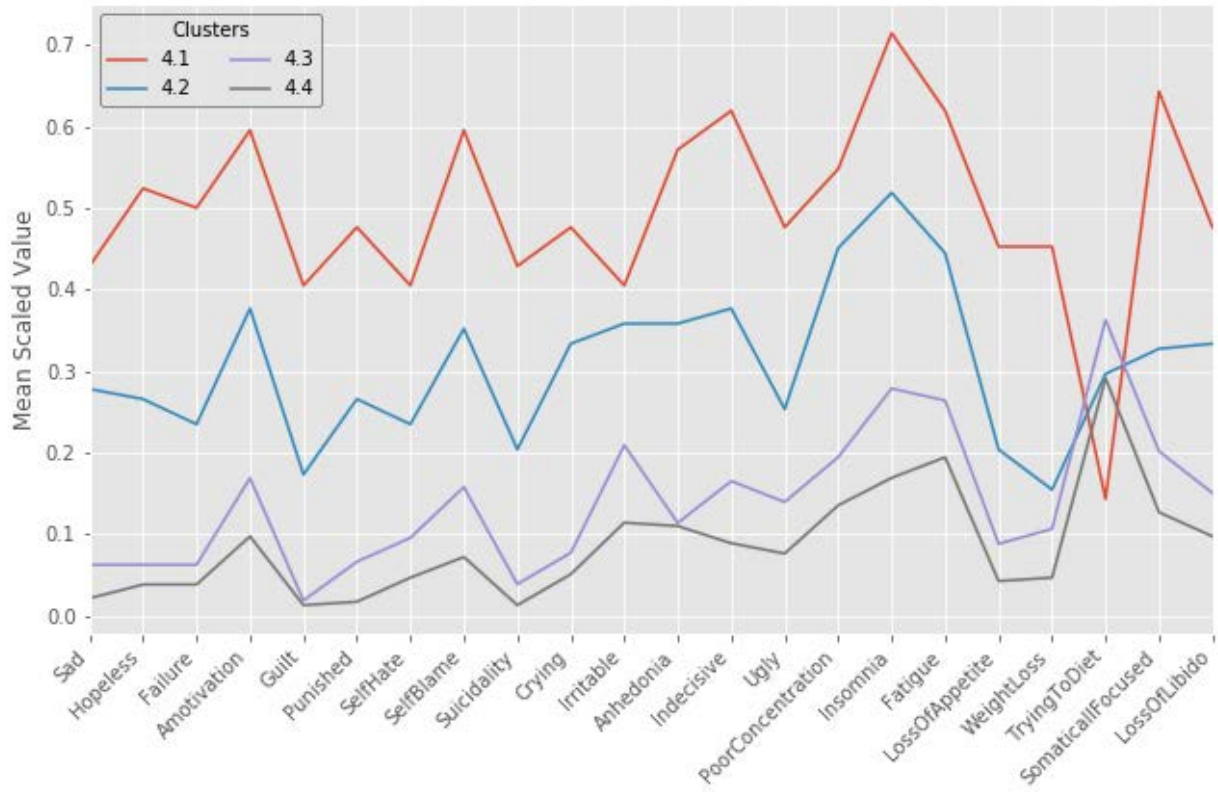

Figure D1.3b. Symptom Profiles for All Items in the Beck Depression Inventory (k=4)

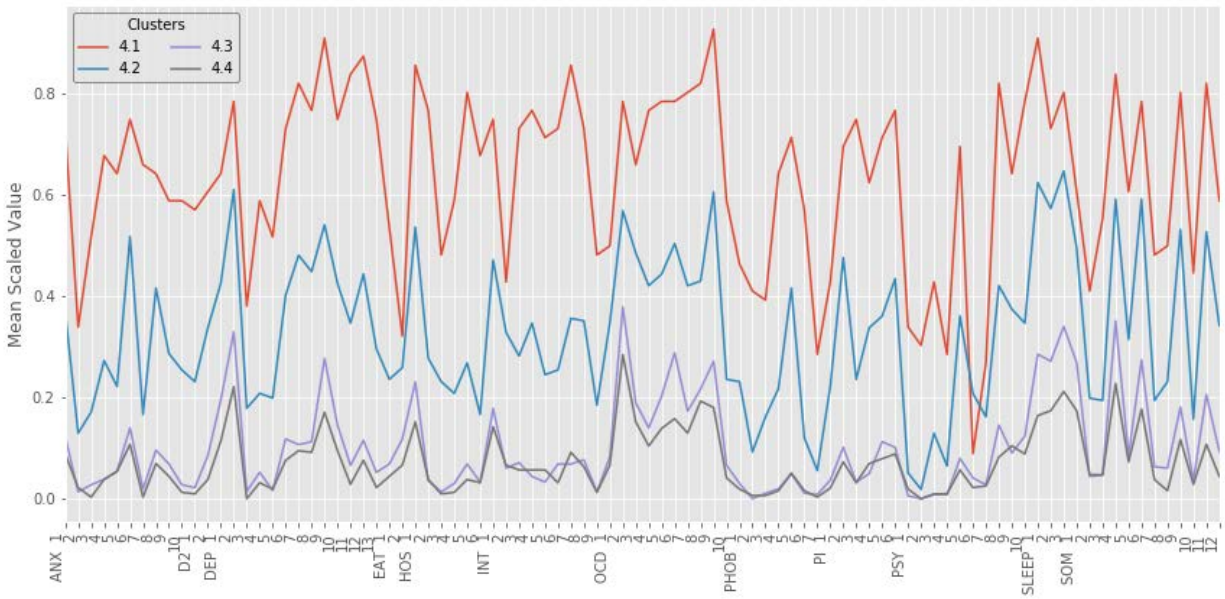

Figure D1.3c. Symptom Profiles for All Items in the Symptom Checklist-90 (k=4)

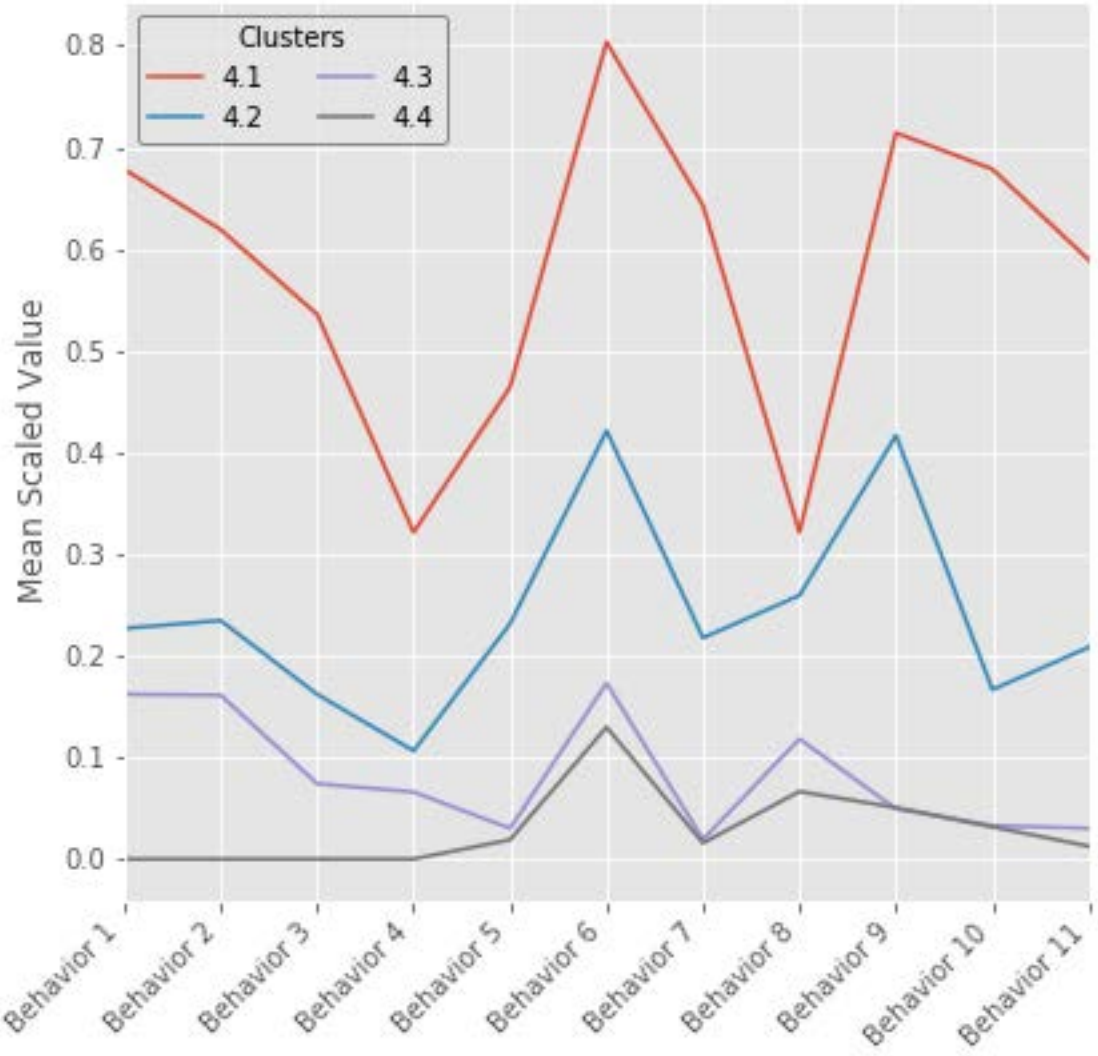

Figure D1.3d. Symptom Profiles for All Behavioral Items (k=4)

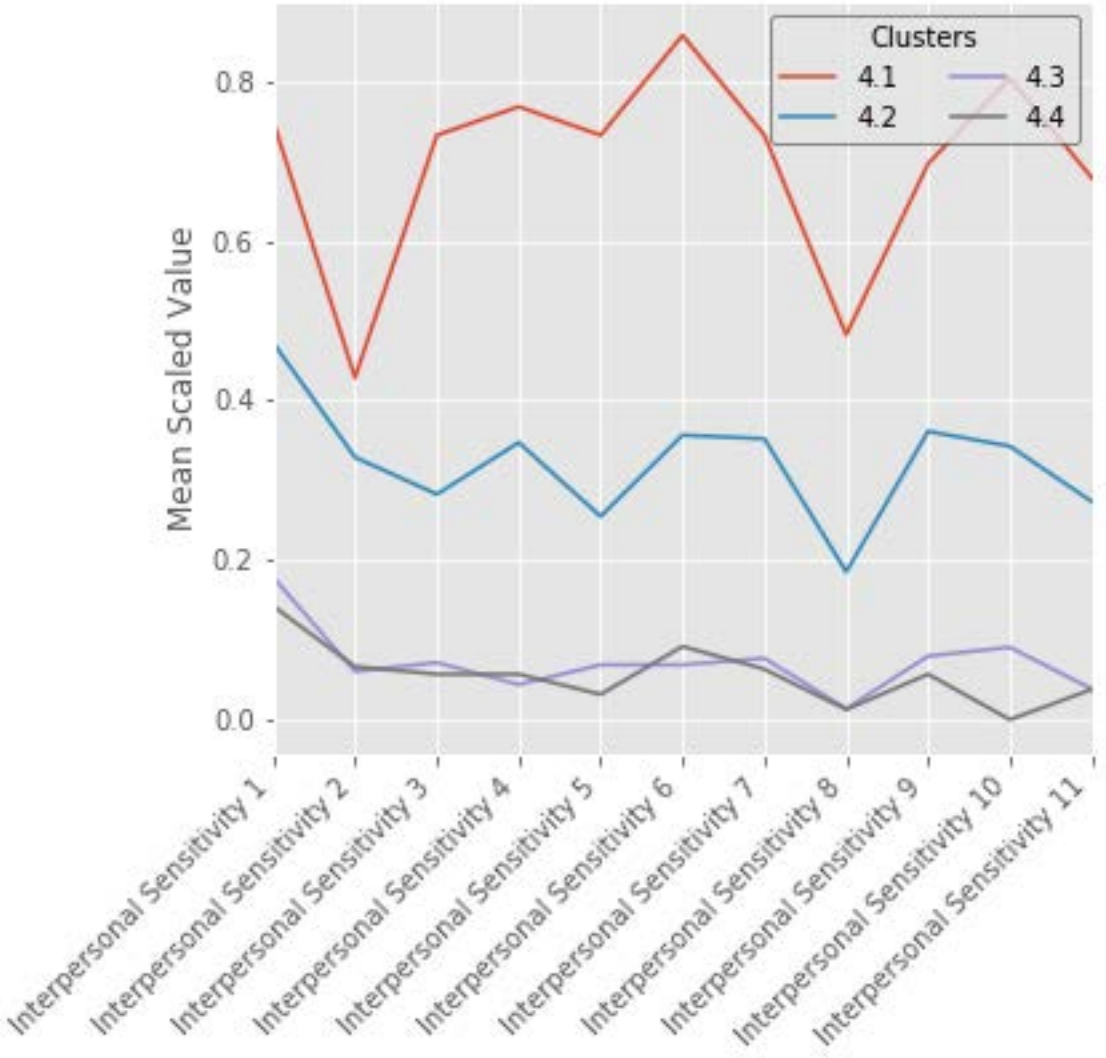

Figure D1.3e. Symptom Profiles for All Interpersonal Sensitivity Items (k=4)

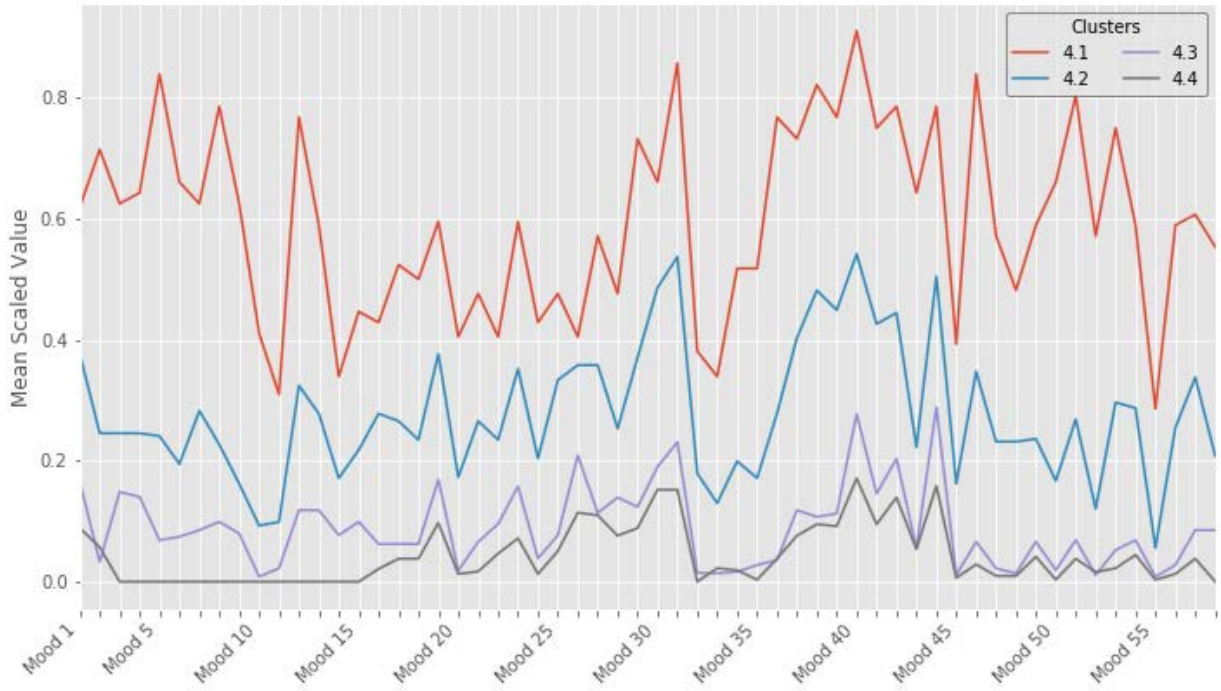

Figure D1.3f. Symptom Profiles for All Mood Items (k=4)

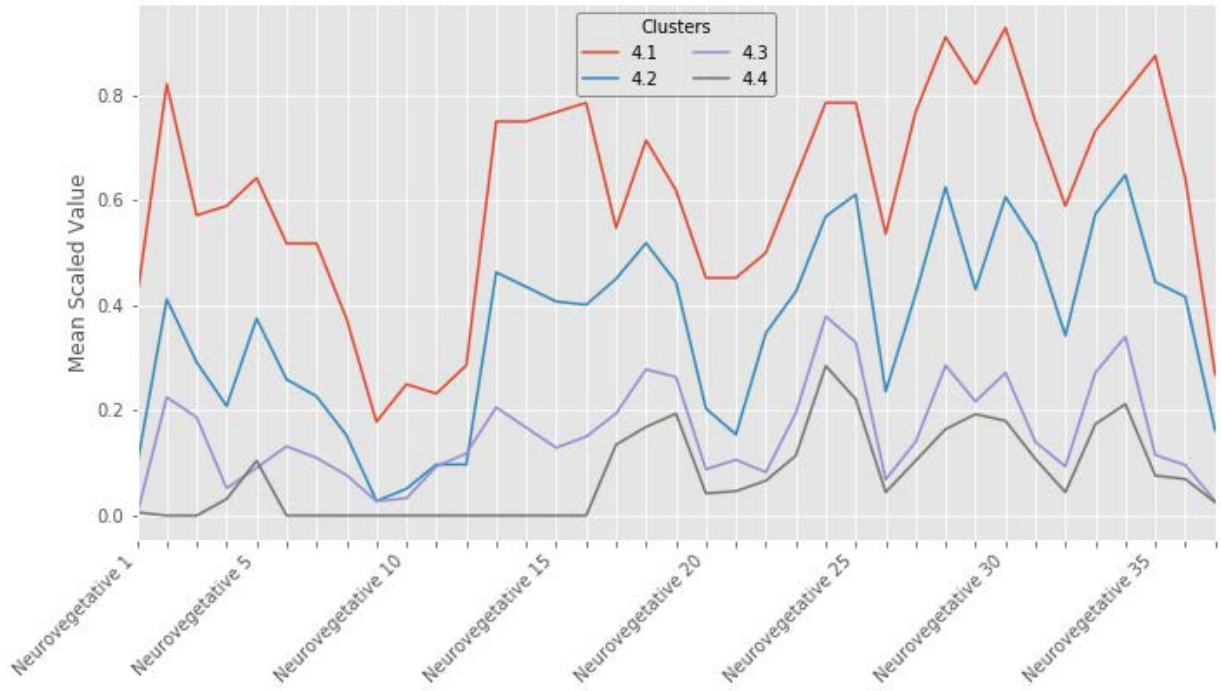

Figure D1.3g. Symptom Profiles for All Neurovegetative Items (k=4)

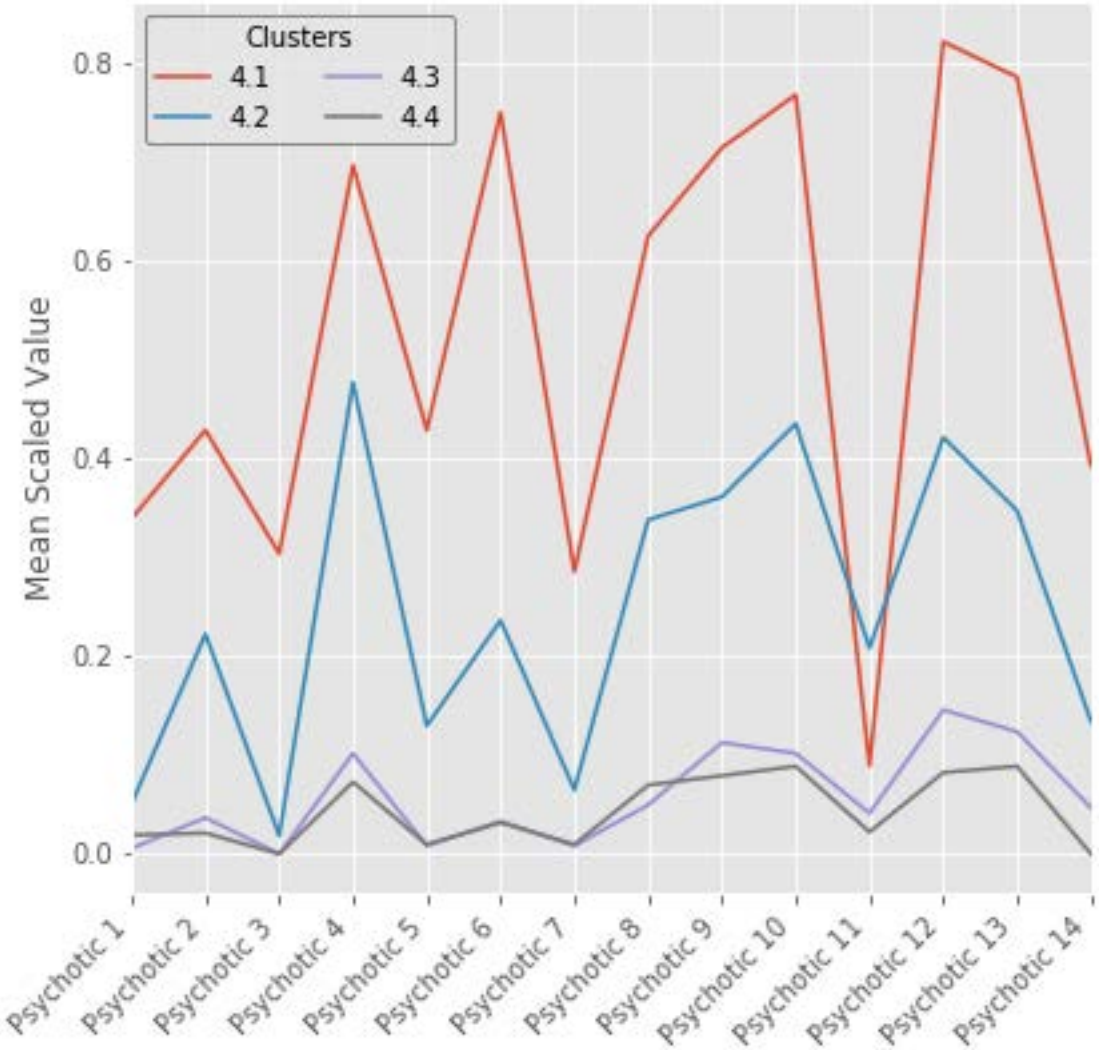

Figure D1.3h. Symptom Profiles for All Psychotic Items (k=4)

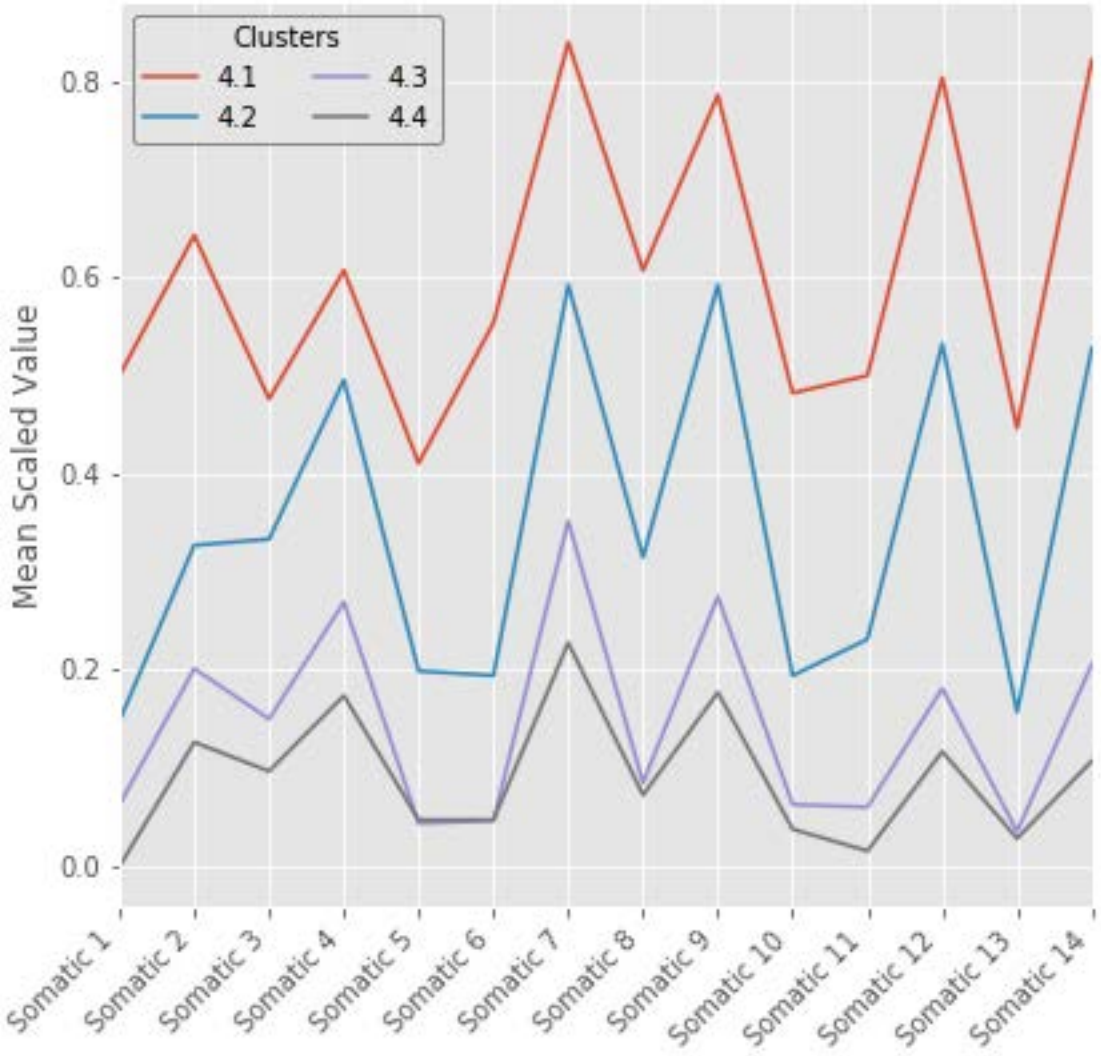

Figure D1.3i. Symptom Profiles for All Somatic Items (k=4)

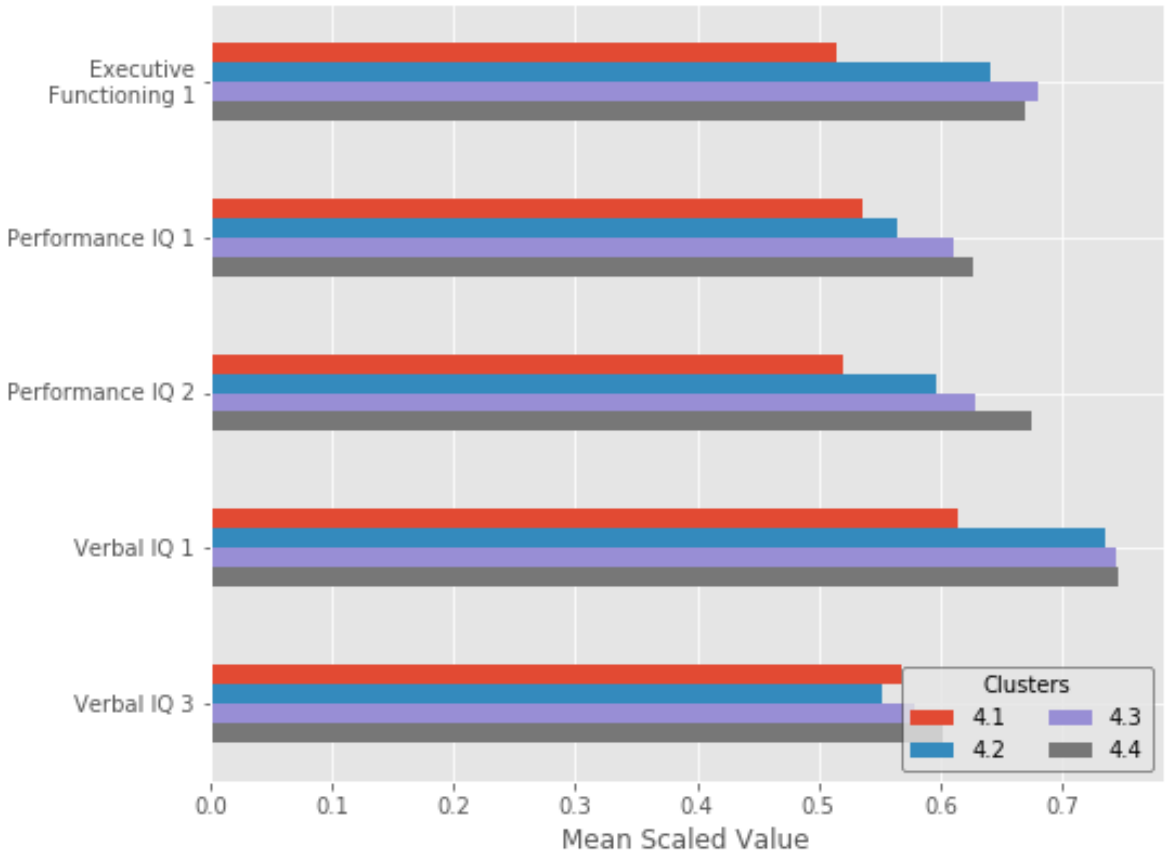

Figure D1.3j. Neuropsychiatric Markers with Significant Differences Across Clusters (k=4)

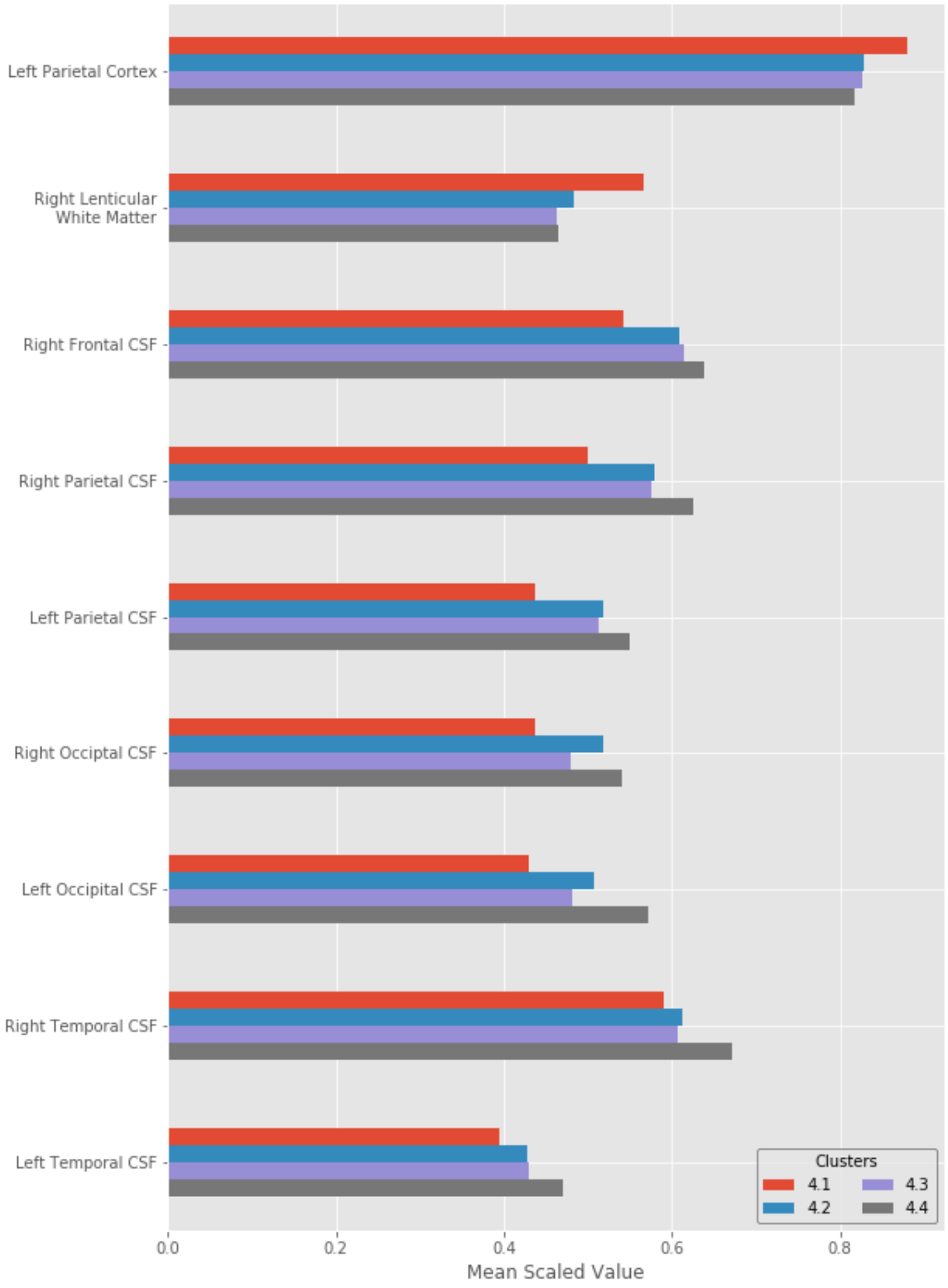

Figure D1.3k. Imaging Markers with Significant Differences Across Clusters (k=4)

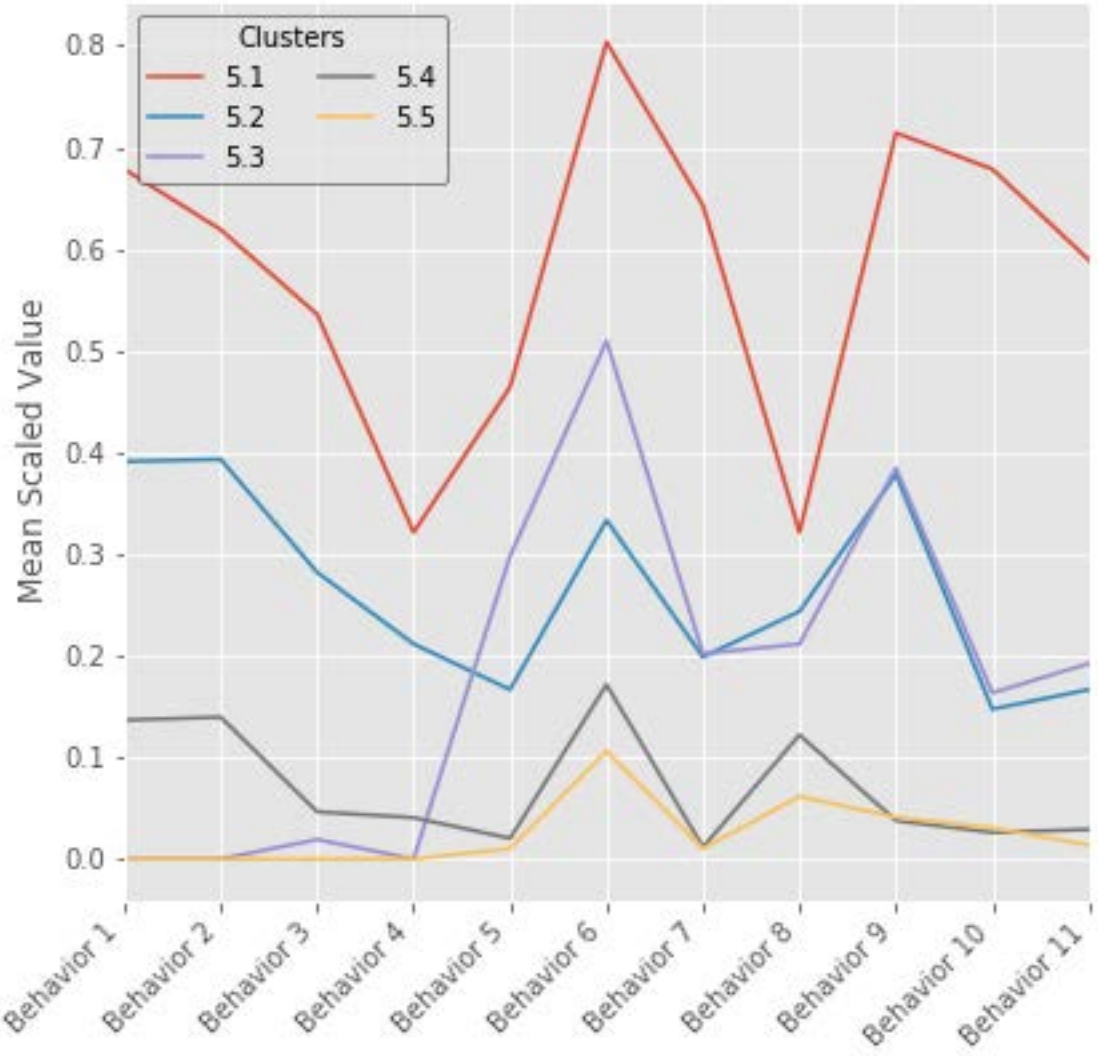

Figure D1.4a. Symptom Profiles for All Behavioral Items (k=5)

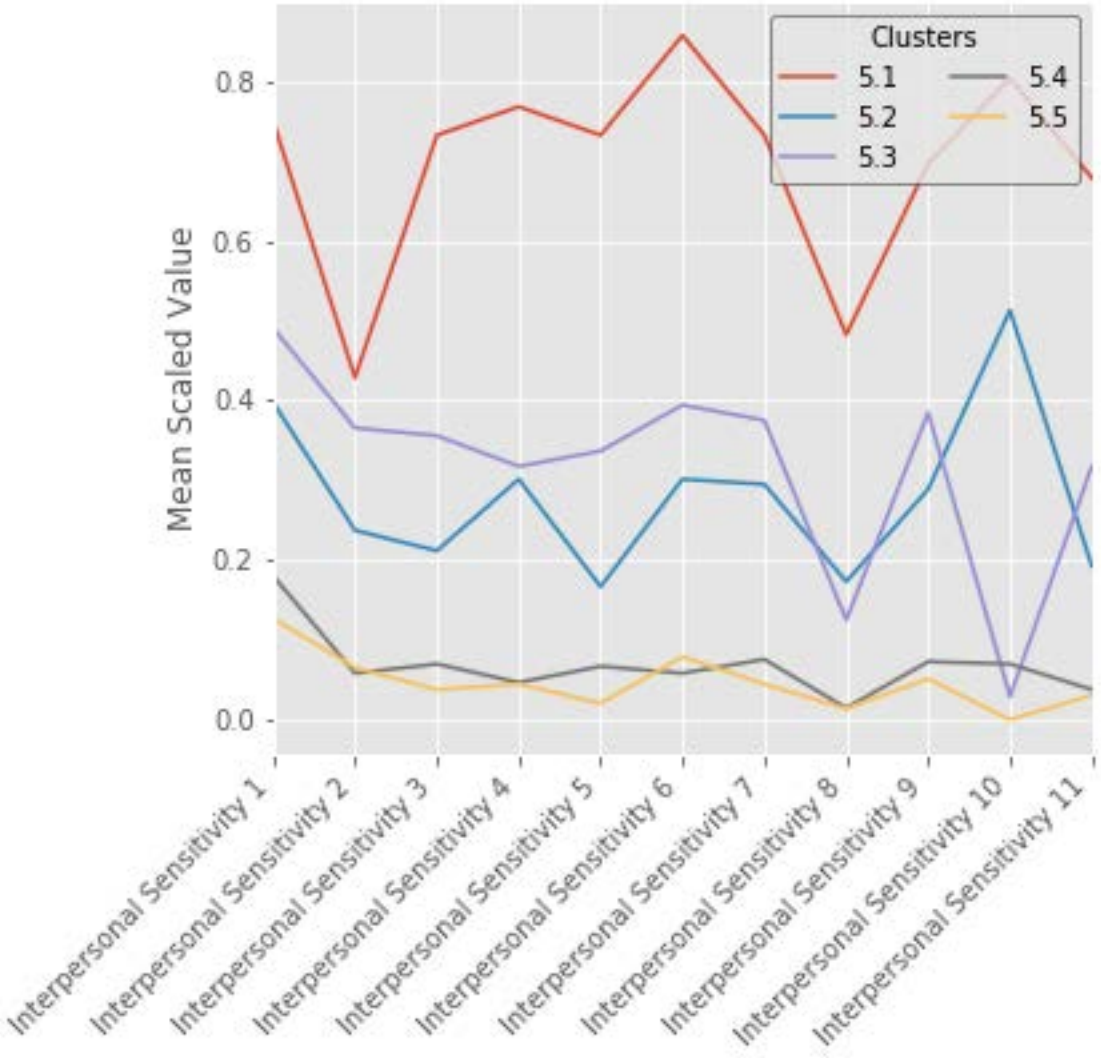

Figure D1.4b. Symptom Profiles for All Interpersonal Sensitivity Items (k=5)

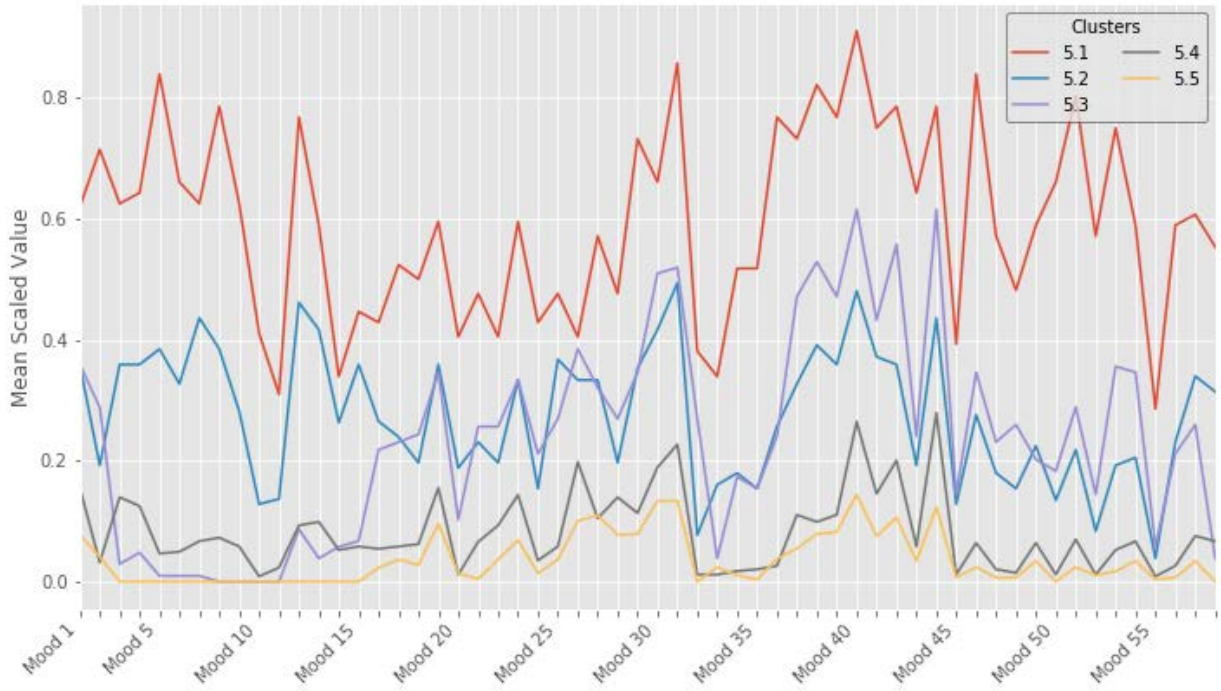

Figure D1.4c. Symptom Profiles for All Mood Items (k=5)

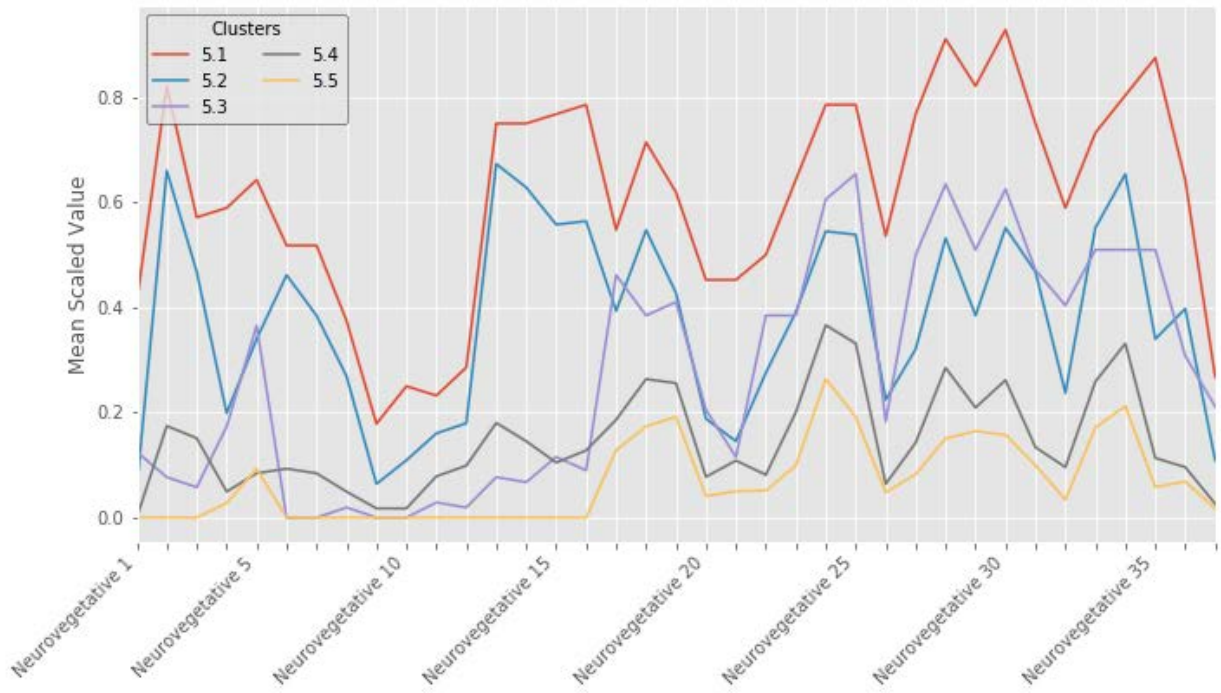

Figure D1.4d. Symptom Profiles for All Neurovegetative Items (k=5)

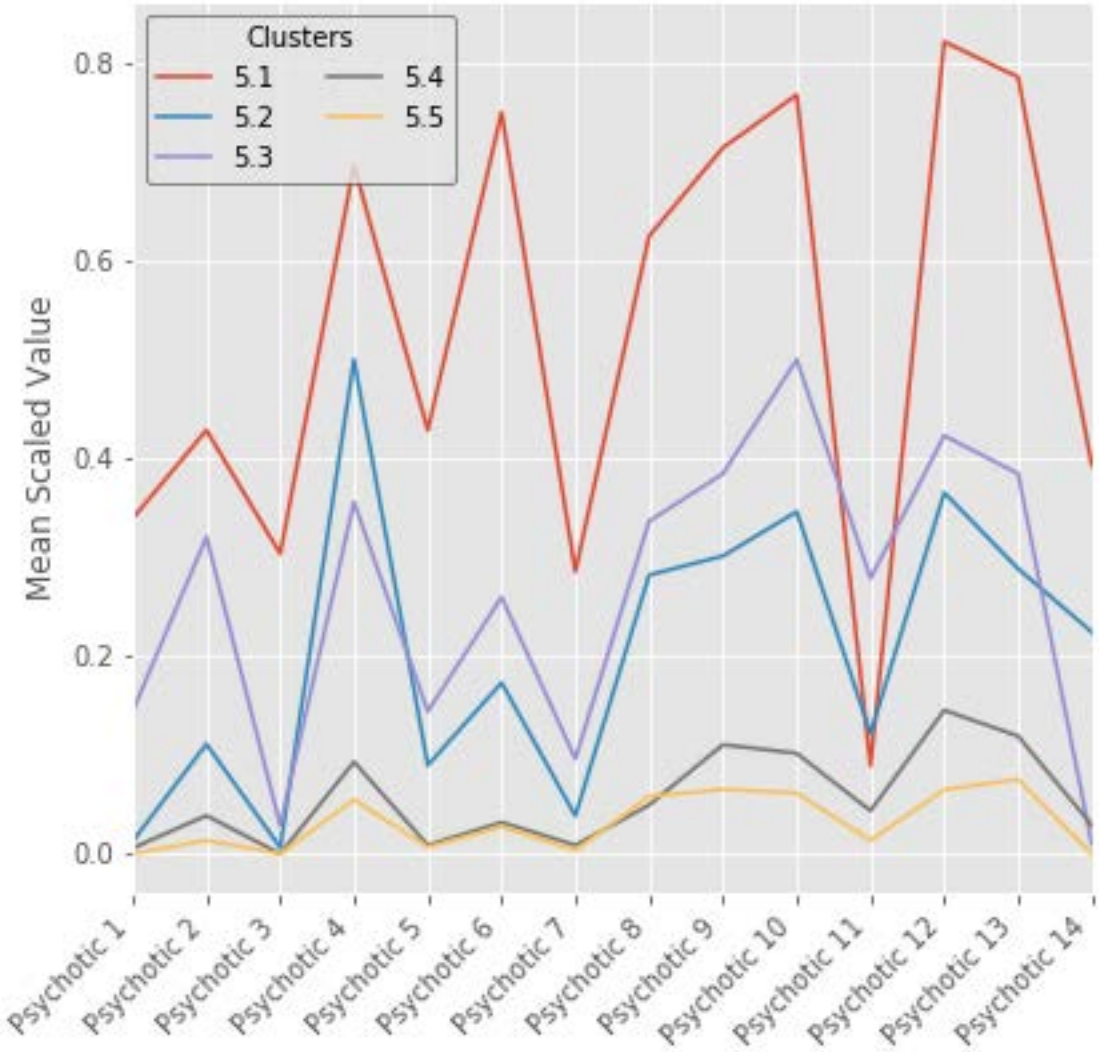

Figure D1.4e. Symptom Profiles for All Psychotic Items (k=5)

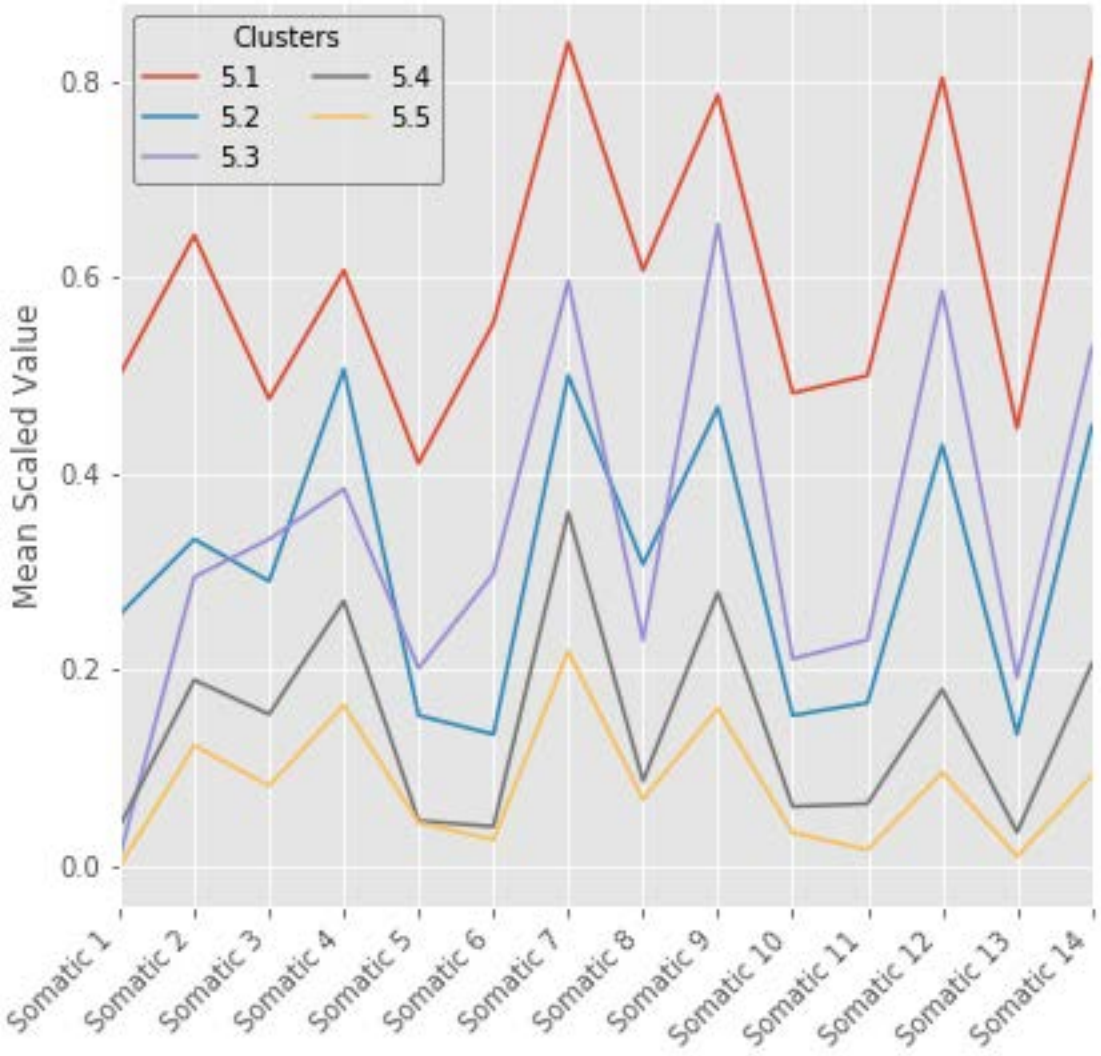

Figure D1.4f. Symptom Profiles for All Somatic Items (k=5)

APPENDIX D2: Cluster Analyses of Gulf War Summary (GS) Dataset for k=2, k=3, k=4, and k-5 Solutions, with Symptom Profiles and Biomarker Differences

This appendix contains the symptom profiles derived using the Gulf War Summary (GS) dataset to cluster individuals, using the full CAPS and Summary Scores from the Ham-D and the SCL-90. As mentioned in the methods, use of the summary scores was necessary to facilitate pooling with the MIRECC dataset in future analyses. As with the GW analyses, for the GS there were 238 males with modified psychiatric symptom data available for use in the cluster analysis, 33 with PTSD diagnoses, 24 with MDD diagnoses, and 60 with alcohol dependence diagnoses. The total number of features used for the GS analysis was 61 (as opposed to 161 in the GW, with 50 items from the CAPS, 10 summary scores from the SCL-90, and one summary item from the HAM-D. The cluster descriptions for the k=2, k=3, k=4, and k=5 analyses are shown in Table D2.1. Biomarker differences across clusters for all four solutions are also presented adjacent to the symptom profiles.

Table D2.1. Details of Cluster Solutions for k=2, k=3, k=4, and k=5

Dataset (# of Clusters) ARI Bootstrap	Cluster (# of Individuals in Cluster)	PTSD	MDD	Other DSM Avg(Range)	Total DSM Avg(Range)	ALC Dependence	$\chi 2$	Childhood Trauma	$\chi 2$
GWS (2)	Cluster2.1 (55)	33	13	0.9(0-4)	2.1(0-7)	24	12.7***	15	2.5
ARI: 0.69	Cluster2.2 (182)	0	11	0.6(0-7)	0.9(0-8)	36		32	
Boot: 0,0									
GWS (3)	Cluster3.1 (14)	12	7	0.7(0-3)	2.9(0-6)	5	12.9**	15	14.8***
ARI: 0.72	Cluster3.2 (70)	20	13	0.8(0-1)	2.4(0-7)	28		25	
Boot: 0,0,0	Cluster3.3 (154)	1	5	0.4(0-7)	1.2(0-15)	27		7	
GWS (4)	Cluster4.1 (12)	11	7	1.1(0-3)	2.9(1-6)	4	26.2****	6	17.5***
ARI: 0.81	Cluster4.2 (45)	0	4	0.9(0-7)	1.1(0-8)	24		10	
Boot: 33,16,0,0	Cluster4.3 (91)	22	6	0.8(0-7)	1.8(0-7)	20		24	
	Cluster4.4 (89)	0	7	0.5(0-3)	0.7(0-4)	12		7	
								6	
GWS (5)	Cluster5.1 (10)	10	6	1.2(0-3)	3.2(1-6)	3	14.2***	5	17.3**
ARI: 0.9	Cluster5.2 (24)	17	3	0.9(0-4)	2.1(0-7)	10		6	
Boot: 5,0,11,2,0	Cluster5.3 (35)	6	5	0.7(0-4)	1.4(0-5)	14		7	
	Cluster5.4 (79)	0	3	0.8(0-7)	1.1(0-8)	21		22	
	Cluster5.5 (89)	0	7	0.5(0-3)	0.7(0-4)	12		7	

GWS: Gulf War Summary
ARI: Adjusted Rand Index between the k-means cluster and hierarchical cluster solutions.
BS: Bootstrap shows the number of times each cluster dissolved during 100 bootstraps of the cluster analyses.
**, p<.01; ***, p<.001; *****, p<.00001

The results of CH and wss measures are shown in Figure D2.1. The CH index revealed that the optimal number of clusters was two, while the wss was at a local minimum between k=3 and k=5, before dropping to another minimum at k=6. The ARI was highest for the five-cluster solution (ARI=0.9), followed by the four- (ARI=0.81), three- (ARI=0.72), and two-cluster (ARI=0.69) solutions. Bootstrapping showed that the k=2, k=3, and k=5 solutions all had clusters that were stable at least 88 percent of the time (Table D2.1).

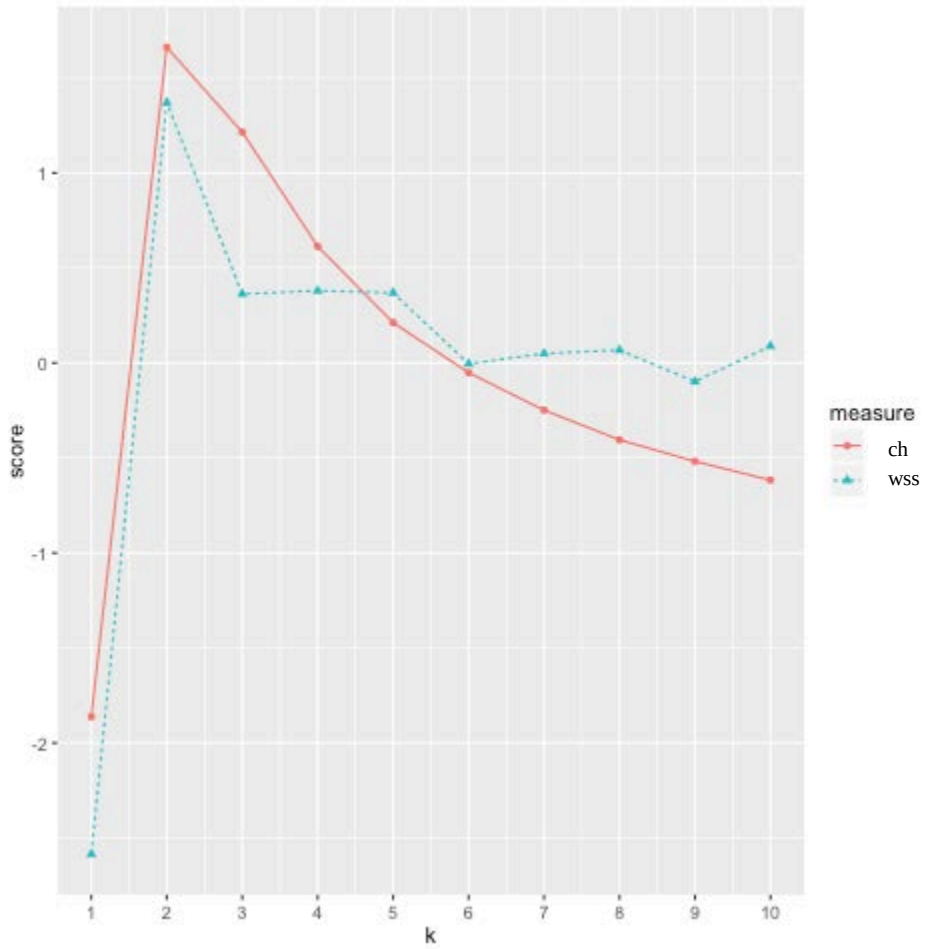

Figure D2.1. Details of Cluster Solutions for k=2, k=3, k=4, and k=5

The k=2 solution partitioned the dataset into two groups. Cluster 2.1 contained all 33 patients with PTSD; these patients made up 75 percent of the cluster. The symptom profiles show that Cluster 2.1 had higher symptom scores across all clinical inventories (Figures D2.2a–D2.2b). The incidence of alcohol dependence was significantly greater in Cluster 2.1 than in Cluster 2.2 ($\chi 2$=12.7, p<.001; Table D2.1).

Cluster 2.1 also performed significantly worse than Cluster 2.2 in four of the neuropsychiatric domains: Executive Functioning (p<0.001), Memory (p<0.001), and Performance IQ (p<.01) and Verbal IQ (p<0.01; Table D2.1, Figure D2.2c). Additionally, there were two imaging markers with significant differences. Cluster 2.1 had smaller average right caudate white matter and right lenticular white matter volumes than Cluster 2.2 (p<0.05). Cluster 2.1 also had smaller average right parietal sulcal CSF volumes (p<0.05) (Table D2.1, Figure D2.2d). Finally, Cluster 2.1 had lower baseline cortisol levels than Cluster 2.2 (p<.01) (Table D2.1, Figure D2.2.e). There were no differences between the two clusters in the genetic markers.

Table. D2.2 Biomarker Differences Across Clusters in the GWS Dataset for k=2

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	ANCOVA P-value	Sig	F	Post-Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=1,227)								
Executive Functioning 1	9.36E+00	3.31E+00	1.09E+01	3.08E+00	1.54E-03	**	10.27	2>1**
Executive Functioning 2	1.04E+01	2.94E+00	1.20E+01	2.68E+00	2.24E-04	***	14.07	2>1***
Memory 1	9.62E+00	3.69E+00	1.14E+01	3.09E+00	6.15E-04	***	12.07	2>1***
Memory 2	1.07E+01	3.34E+00	1.22E+01	2.78E+00	1.17E-03	**	10.8	2>1**
Performance IQ 2	3.85E+01	1.24E+01	4.34E+01	1.18E+01	4.71E-03	**	8.15	2>1**
Imaging Volumes (D.F.=1,164)								
Right Caudate White Matter	8.51E-04	4.43E-04	7.22E-04	2.38E-04	1.69E-02	*	6.83	1>2*
Right Lenticular White Matter	2.42E-03	5.92E-04	2.22E-03	4.44E-04	2.40E-02	*	5.19	1>2*
Right Parietal CSF	2.08E-02	4.70E-03	2.29E-02	5.00E-03	1.15E-02	*	6.54	2>1*
Left Temporal CSF	1.29E-02	2.34E-03	1.39E-02	3.11E-03	2.93E-02	*	4.42	N.S.
Cortisol Measures (D.F.=1,150)								
Baseline AM Cortisol Day 1 of DST	3.07E+00	6.70E-01	3.37E+00	4.90E-01	4.17E-03	**	6.26	2>1**

D.F.: degrees of freedom
S.D.: standard deviation
N.S.: not significant
TK: Tukey-Kramer
*, p<.05; **, p<.01; ***, p<.001

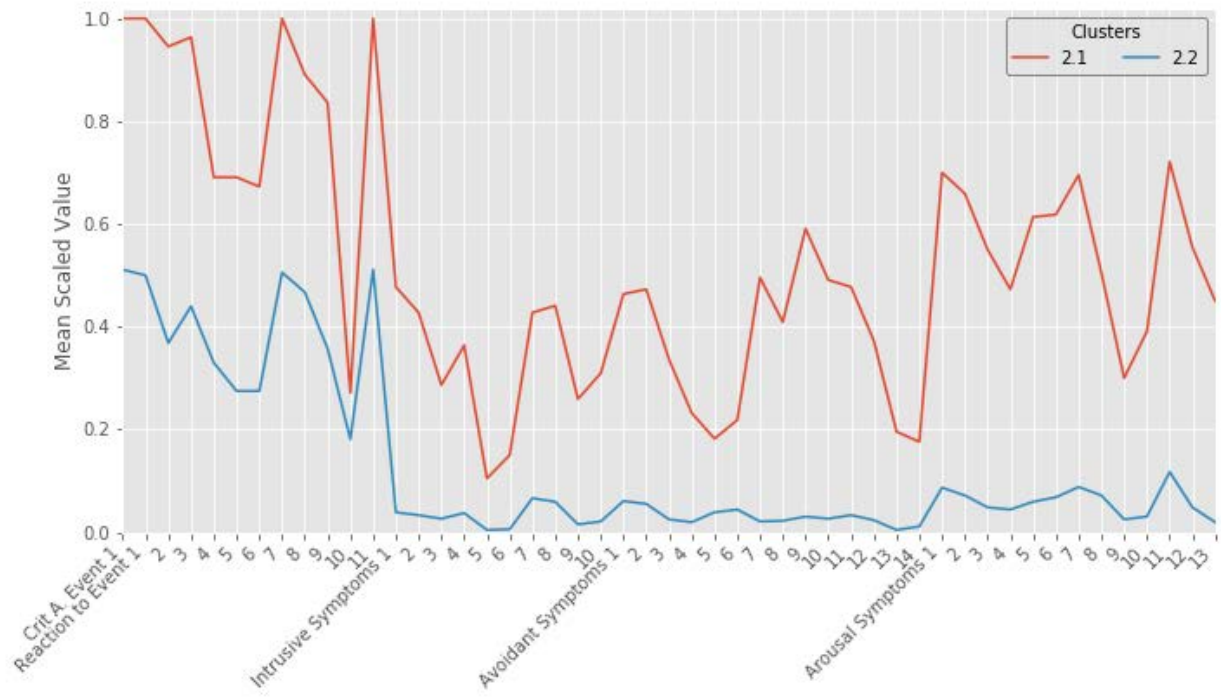

Figure D2.2a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=2)

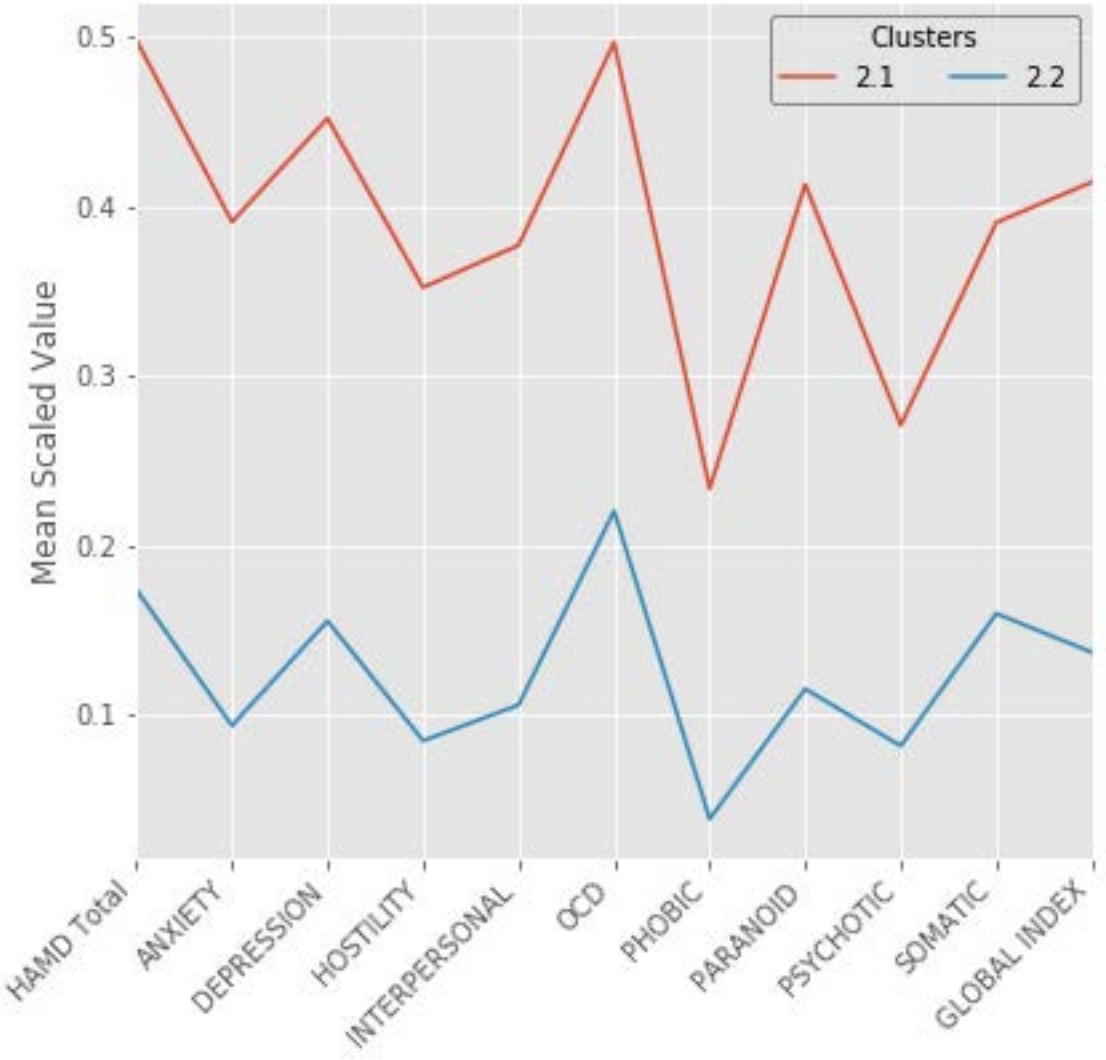

Figure D2.2b. Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 (k=2)

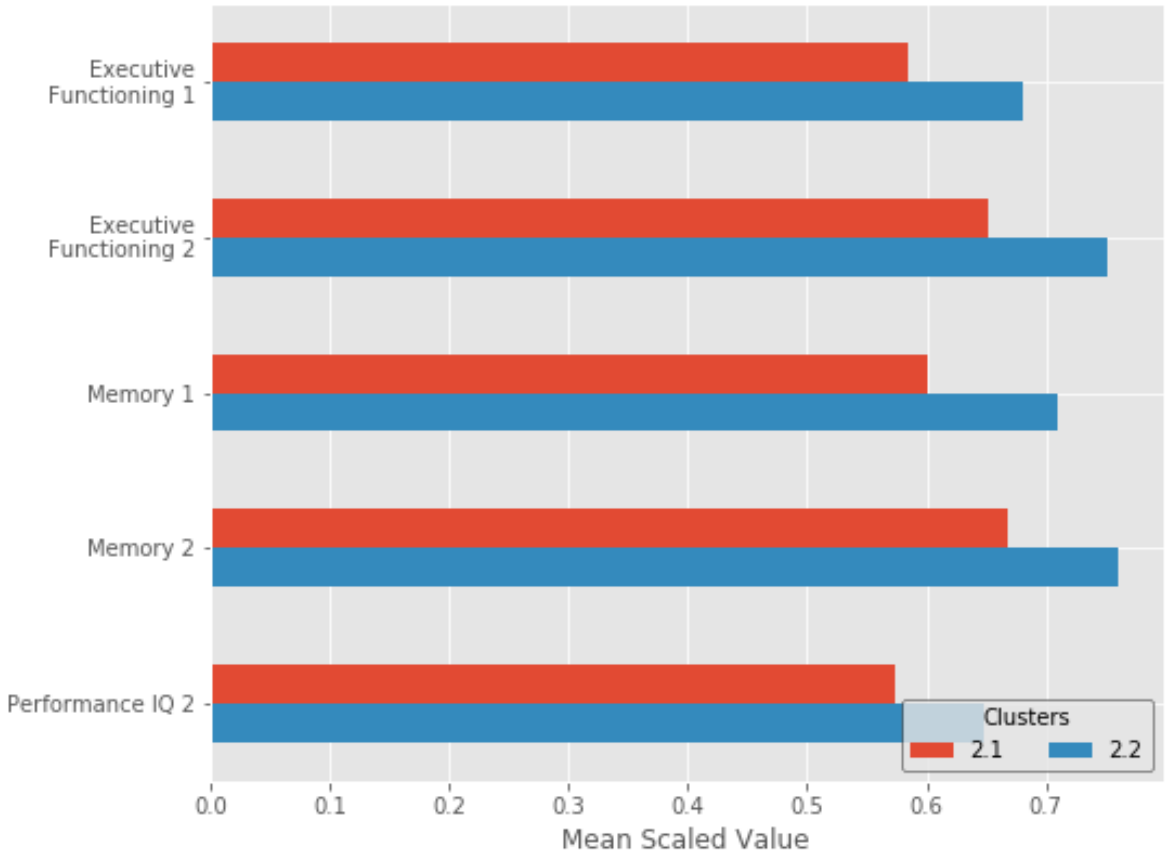

Figure D2.2c. Neuropsychiatric Markers with Significant Differences Across Clusters (k=2)

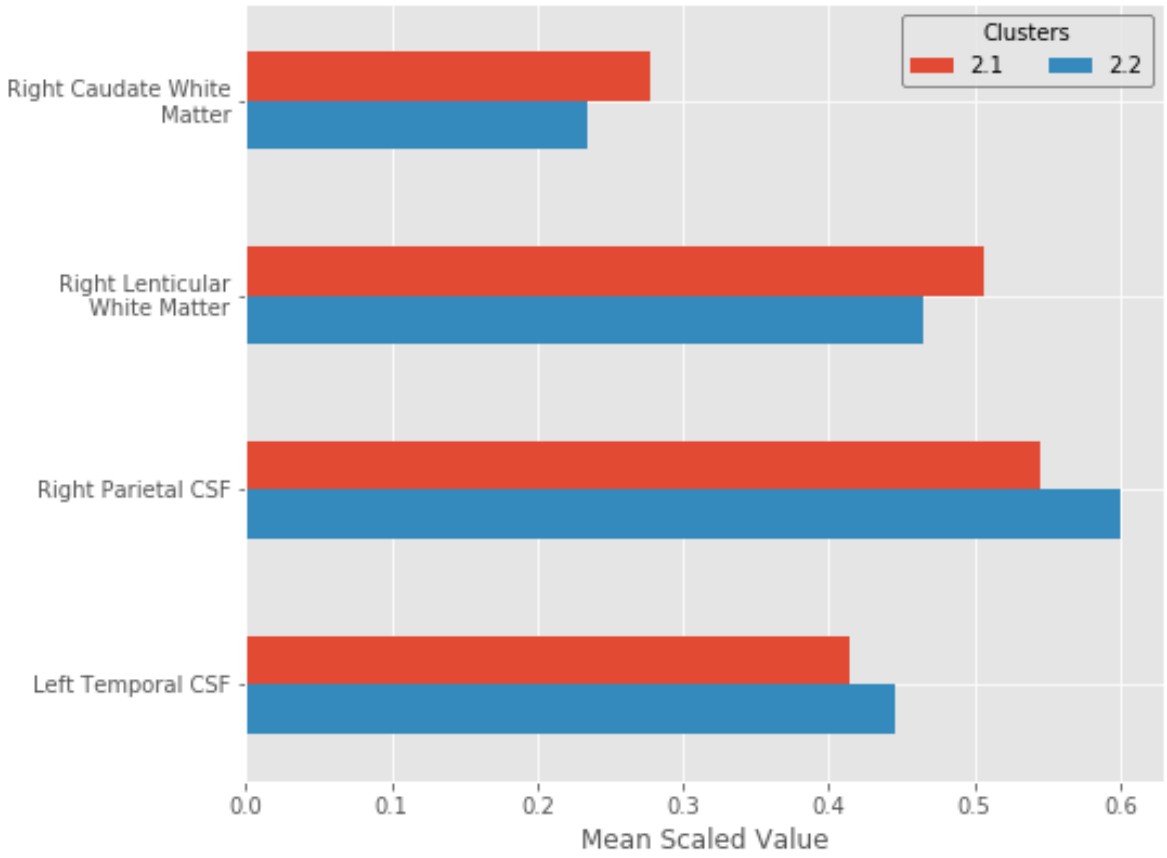

Figure D2.2d. Imaging Markers with Significant Differences Across Clusters (k=2)

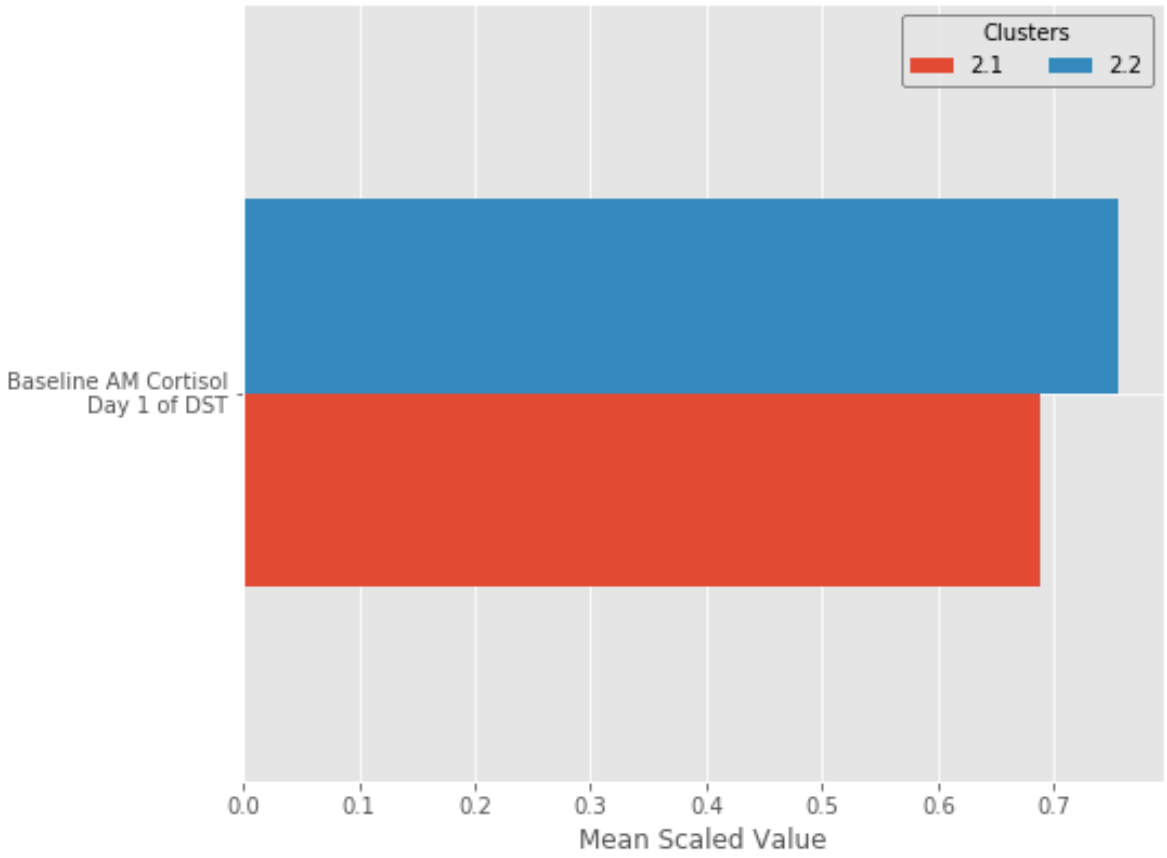

Figure D2.2e. Endocrine Markers with Significant Differences Across Clusters (k=2)

In the three-cluster solution, 32 out of 33 of the individuals with PTSD remained in Cluster 3.1, representing 73 percent of this group. Cluster 3.2 comprised the remaining individuals with a PTSD diagnosis. Cluster 3.1 had, on average, 2.2 DSM diagnoses per individual, compared with 1.2 diagnoses per individual in Cluster 3.2 and 0.7 in Cluster 3.3 (Table D2.1). As expected, symptom profiles from the CAPS clinical inventory were highest in Cluster 3.1, followed by Clusters 3.2 and 3.3 (Figure D2.3a). Symptoms included in the Ham-D and SCL-90 summary scores were also highest in Cluster 3.1 (Figure D2.3b). Generally, these symptom profiles tended to be minimally higher in Cluster 3.2 than Cluster 3.3 with the exceptions of Cluster 3.2 having slightly higher mean scores in the phobic and interpersonal summary scores (Figure 6.3b). The incidence of alcohol dependence across all three clusters differed significantly ($\chi 2=12.9$, $p<.001$), with the percentage of individuals with alcohol dependence being the greatest in 3.1 and least in 3.3, decreasing as the symptom profiles lowered. This decreasing pattern as the cluster symptom profile decreased was also seen in the incidence of childhood trauma ($\chi 2=14.8$, $p<.001$; Table D2.1).

Cluster 3.1 had lower neuropsychiatric scores than Cluster 3.2, Cluster 3.3, or both clusters combined, in the Executive Functioning ($p<.05$, $p<.01$, $p<.001$), Memory ($p<.05$, $p<.01$, $p<.001$), and Performance IQ domains ($p<.05$, $p<.001$) (Table D2.3, Figure 6.3c). Cluster 3.1 also had larger average right caudate white matter than Cluster 3.2 ($p<0.05$), and smaller volumes of right temporal ($p<0.001$) and left temporal CSF ($p<0.01$) than Cluster 3.3 (Table D2.2, Figure D2.3d). Furthermore, the cluster with a single PTSD patient, Cluster 3.3, had smaller average bilateral occipital cortex volumes ($p<.05$) and larger average total subcortical CSF volumes (as well as individually larger CSF volumes in the right parietal, bilateral occipital,

and bilateral temporal lobes ($p<.05$, $p<.010$) than Cluster 3.2 (Table D2.3, Figure D2.3d). There were no significant differences found in cortisol measures between any of the clusters.

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	Cluster 3 Mean	S.D.	ANCOVA P-value	Sig	F	Post Hoc Cluster Difference
Neuropsychiatric (D.F.=2,226)										
Executive Functioning 1	8.98E+00	3.30E+00	1.10E+01	3.04E+00	1.08E+01	3.12E+00	1.01E-03	**	7.12	2>1**,3>1**
Executive Functioning 2	1.02E+01	2.94E+00	1.22E+01	2.41E+00	1.17E+01	2.97E+00	1.82E-04	***	8.95	2>1***,3>1**
Learning 3	4.70E+01	1.02E+01	5.16E+01	9.67E+00	5.06E+01	1.02E+01	3.27E-02	*	3.47	2>1*
Memory 1	9.12E+00	3.70E+00	1.15E+01	2.99E+00	1.13E+01	3.19E+00	1.53E-04	***	9.13	2>1***,3>1**
Memory 2	1.03E+01	3.38E+00	1.23E+01	2.71E+00	1.20E+01	2.86E+00	6.90E-04	***	7.52	2>1***,3>1**
Memory 4	2.21E+01	7.87E+00	2.49E+01	7.50E+00	2.58E+01	8.15E+00	3.97E-02	*	3.27	3>1*
Performance IQ 2	3.67E+01	1.17E+01	4.26E+01	1.20E+01	4.48E+01	1.17E+01	5.52E-04	***	7.76	2>1*,3>1***
Imaging Volumes (D.F.=2,163)										
Right Occipital Cortex	1.88E-02	2.10E-03	1.93E-02	2.00E-03	1.83E-02	2.20E-03	1.37E-02	*	4.4	2>3*
Left Occipital Cortex	1.93E-02	2.50E-03	1.96E-02	2.10E-03	1.85E-02	2.30E-03	1.18E-02	*	4.56	2>3*
Right Caudate White Matter	8.81E-04	4.73E-04	7.00E-04	2.12E-04	7.43E-04	2.57E-04	1.31E-02	*	4.45	1>2*
Right Parietal CSF	2.08E-02	4.70E-03	2.18E-02	4.90E-03	2.40E-02	4.80E-03	1.69E-03	**	6.64	3>1**,3>2*
Left Parietal CSF	2.12E-02	4.51E-03	2.29E-02	4.56E-03	2.05E-02	4.78E-03	1.04E-02	*	4.69	3>1*
Right Occipital CSF	5.23E-03	1.50E-03	4.92E-03	1.20E-03	5.53E-03	1.40E-03	3.23E-02	*	3.51	3>2*
Left Occipital CSF	5.01E-03	1.30E-03	4.71E-03	1.10E-03	5.53E-03	1.50E-03	6.70E-04	***	7.65	3>2*
Right Temporal CSF	1.43E-02	2.60E-03	1.45E-02	2.30E-03	1.57E-02	2.90E-03	1.49E-03	**	6.78	3>2**,3>1**
Subcortical CSF	1.00E-03	2.89E-04	9.58E-04	1.99E-04	1.08E-03	2.33E-04	1.14E-02	*	4.6	2>1**
Endocrine Measures (D.F.=2,149)										
Baseline AM Cortisol Day 1 of DST	3.22E+00	6.02E-01	3.22E+00	6.02E-01	3.44E+00	4.65E-01	4.37E-02	*	3.2	N.S.

D.F., degrees of freedom
S.D., standard deviation
N.S., not significant
TK, Tukey-Kramer
*, p<.05; **, p<.01; ***, p<.001

Table D2.3 Biomarker Differences Across Clusters in the GWS Dataset for k=3

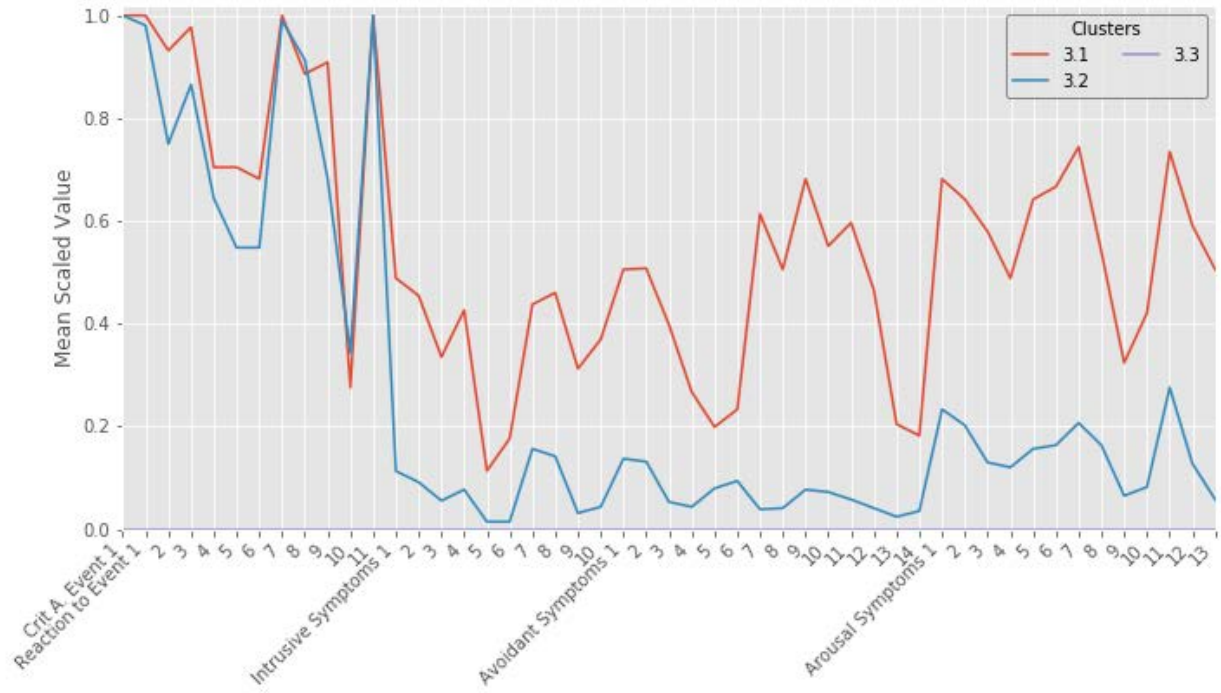

Figure D2.3a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=3)

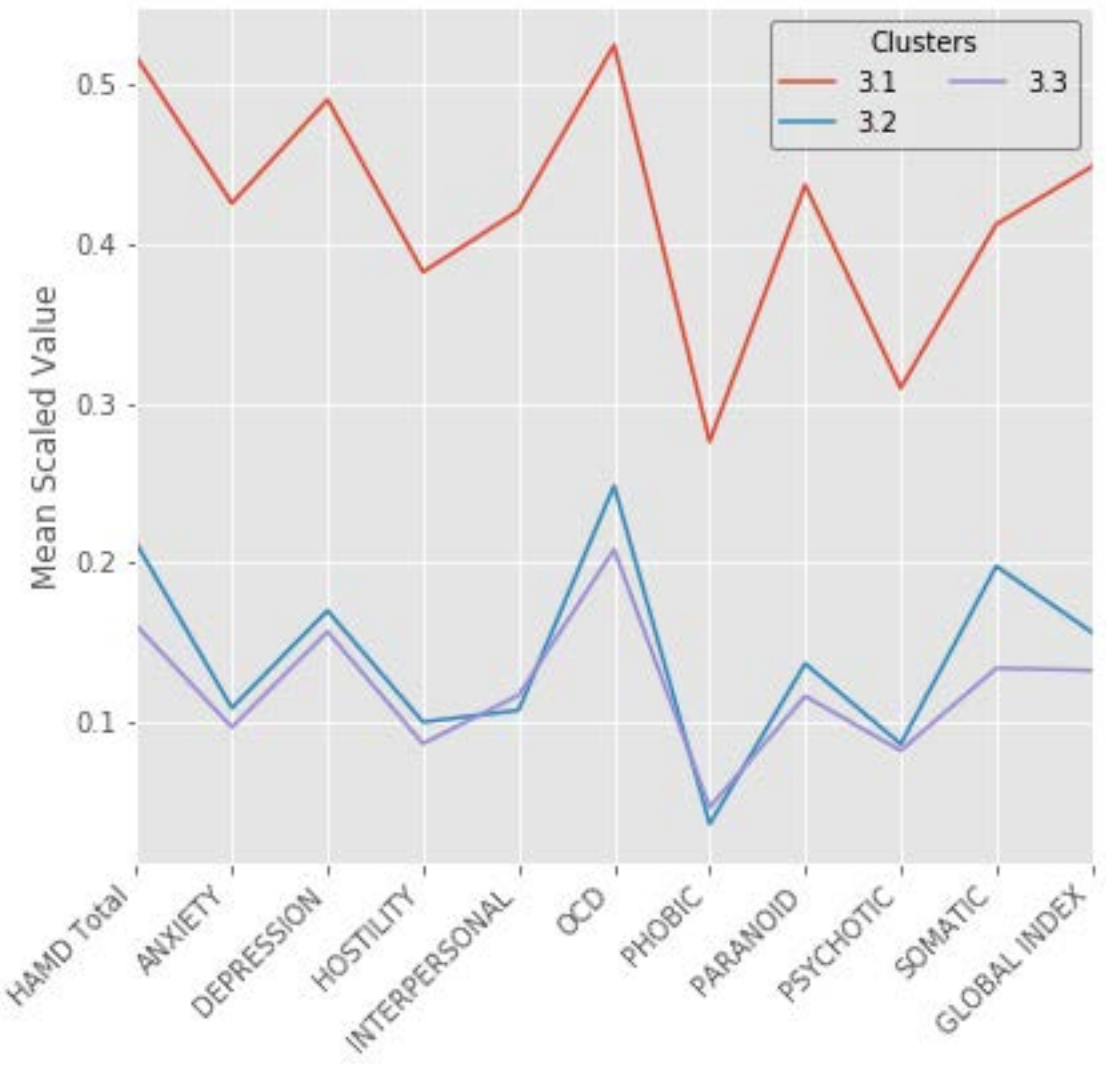

Figure D2.3b. Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 (k=3)

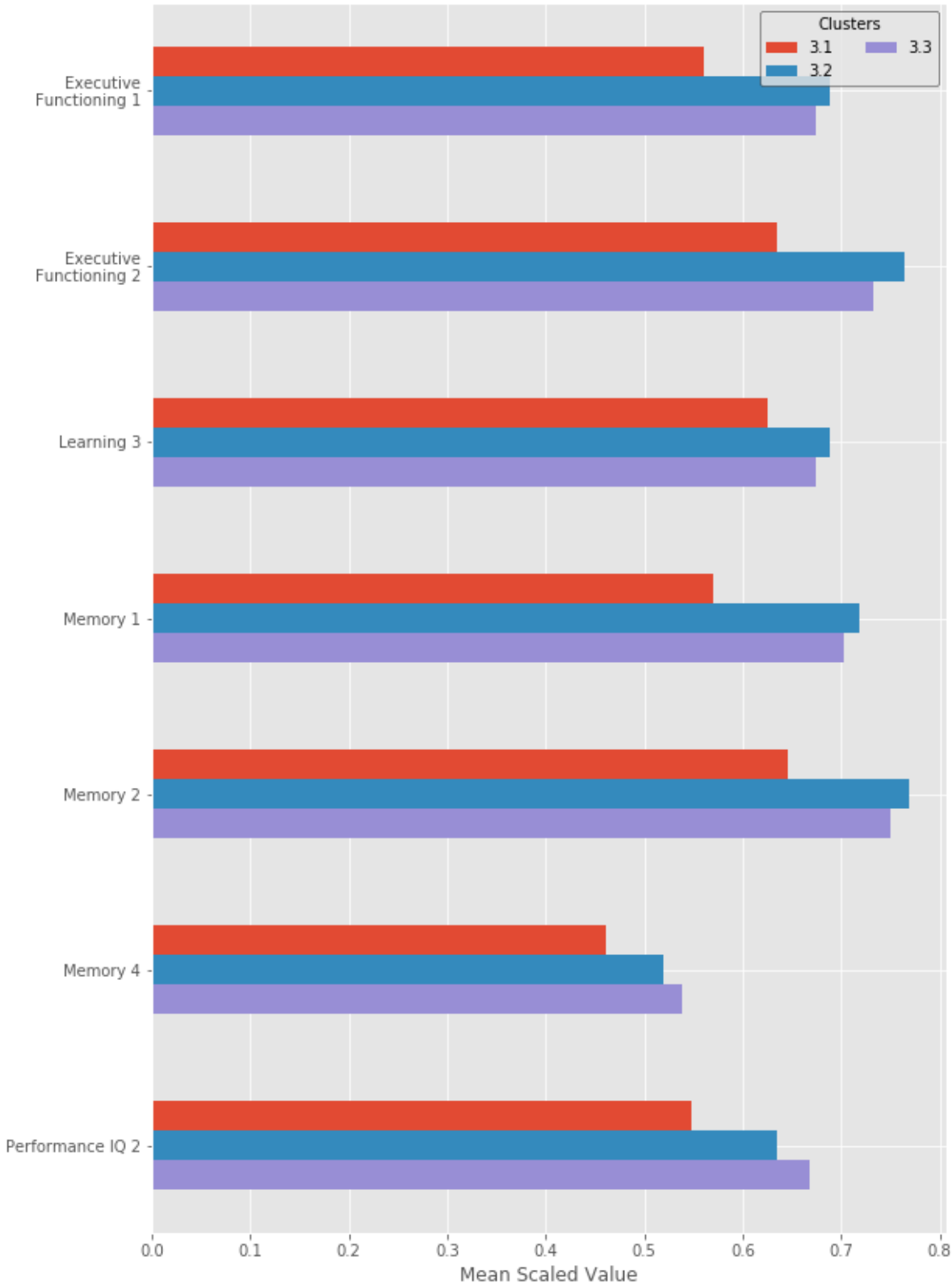

Figure D2.3c. Neuropsychiatric Markers with Significant Differences Across Clusters (k=3)

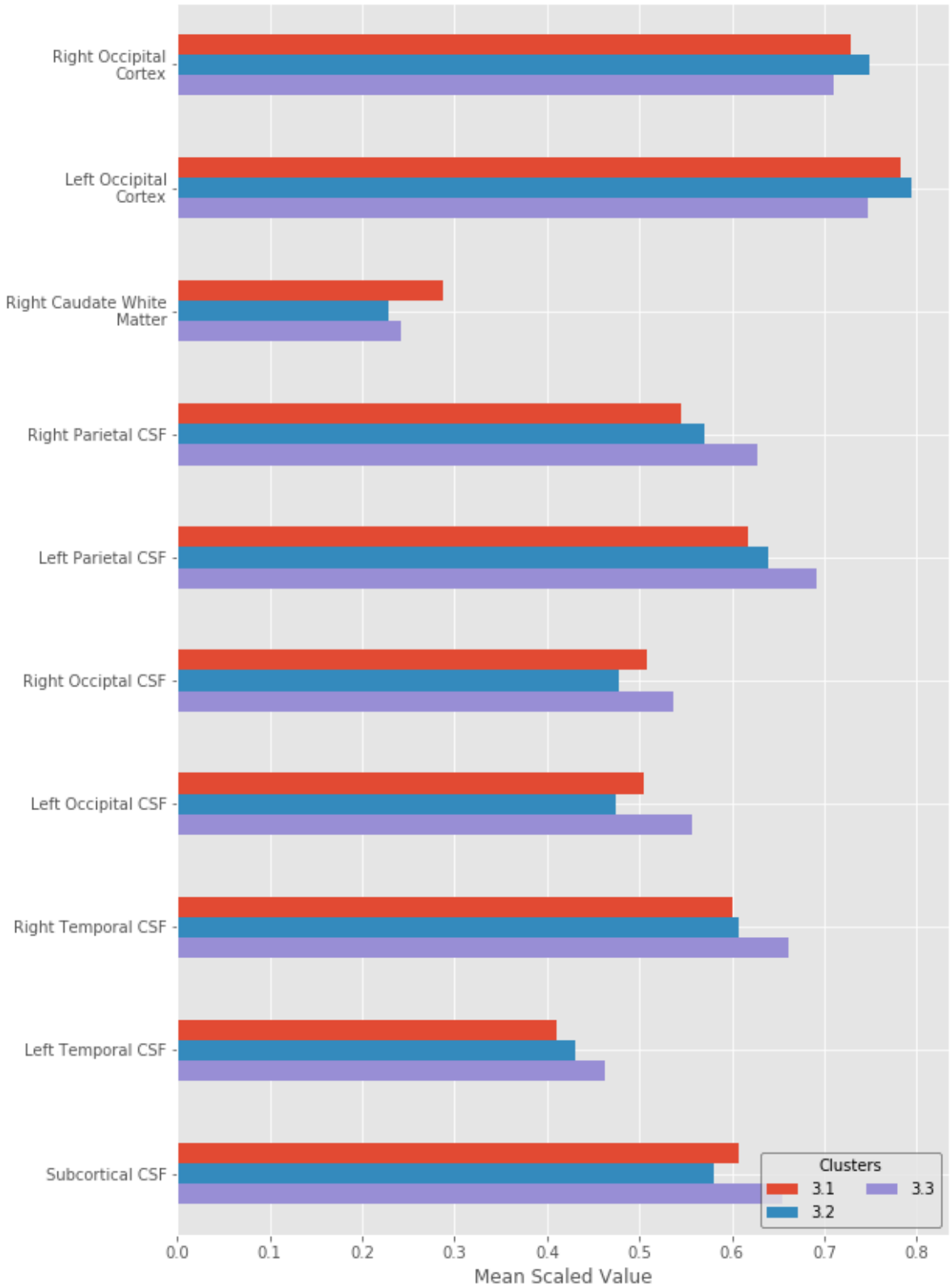

Figure D2.3d. Imaging Markers with Significant Differences Across Clusters (k=3)

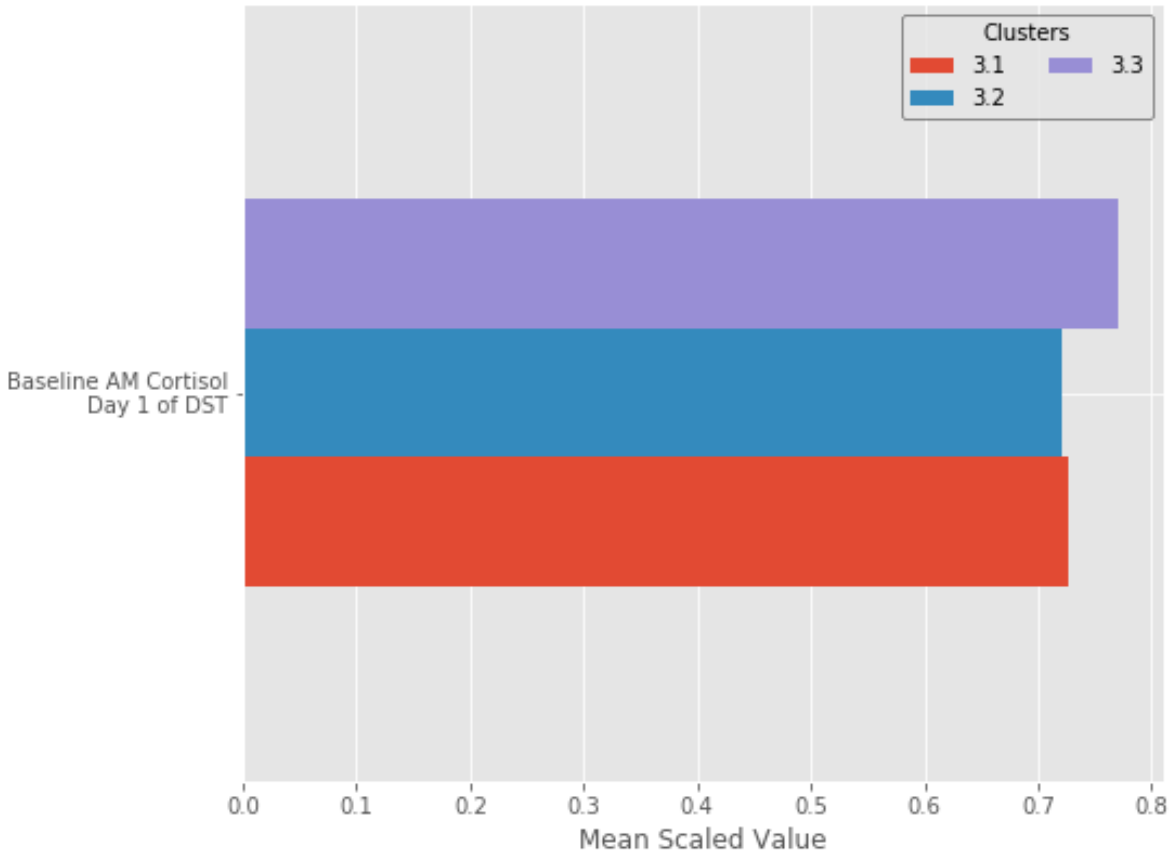

Figure D2.3e. Endocrine Markers with Significant Differences Across Clusters (k=3)

In the four-cluster solution, 67 percent of the individuals with a diagnosis of PTSD were placed in Cluster 4.3; these patients represented 24 percent (22/91) of the cluster. The remaining 33 percent of PTSD patients represented 92 percent (11/12) of Cluster 4.1. Of note, Cluster 4.1 dissolved in the bootstrapping analyses 33 percent of the time. The mean number of DSM diagnoses for Cluster 4.3 was 1.8, while the same value for Cluster 4.1 was 2.9 (Table D2.1). Cluster 4.1, which had a higher average DSM diagnosis per cluster, also had the highest symptom profile across the CAPS, Ham-D, and SCL-90 sub-scores (Figures D2.4a and D2.4b). Surprisingly, the next highest symptom profile was in Cluster 4.2, the cluster with the greatest number of individuals diagnosed with alcohol dependence (24/70), composing 53 percent of this group. This finding was surprising because none of the individuals in Cluster 4.2 carried a PTSD diagnosis, but they still reported higher average scores in CAPS than the average individual in Cluster 4.3, where 24 percent of the individuals had a PTSD diagnosis. This may be explained by the fact that several individuals in Cluster 4.2 had Criterion A experiences without having intense levels of symptoms. In the Ham-D symptom measure and the OCD, somatic, and general SCL-90 summary scores, Cluster 4.1 continued have higher average scores than Clusters 4.2, 4.3, and 4.4 (in descending order). However, Clusters 4.3 and 4.4 had very similar summary scores (Figure D2.4b). The incidence of alcohol dependence across all four clusters differed significantly ($\chi 2=26.2$, $p<.00001$; Table D2.1), with the percentage of individuals with alcohol dependence being the greatest in 4.2, followed by 4.1, 4.3, and 4.4. A decreasing pattern as the cluster symptom profile decreased was seen in the incidence of childhood trauma ($\chi 2=7.5$, $p<.001$; Table D2.1), which was highest in Cluster 4.1 and lowest in Cluster 4.4.

Cluster 4.1 performed worse than Clusters 4.4, 4.3, and 4.2 in two neuropsychiatric domains: Executive Functioning ($p<.05$) and Verbal IQ ($p<.05$, $p<.01$, $p<.001$), worse than

Clusters 4.4 and 4.3 in Memory ($p<.05$, $p<.01$), and worse than Cluster 4.4 in Performance IQ ($p<.001$; Table D2.4, Figure D2.4.c). Interestingly, Cluster 4.2, the cluster with the highest number and percentage of individuals with alcohol dependence, scored worse than Cluster 4.4 in Executive Functioning ($p<0.05$), and worse than Cluster 4.3 in Executive Functioning ($p<0.01$) and Memory ($p<0.01$) (Table D2.4, Figure D2.4c). Cluster 4.2, the cluster with the highest number of alcohol-dependent individuals, had significantly reduced right parietal CSF volumes ($p<0.05$) and baseline cortisol ($p<0.05$) than Cluster 4.4 (Table D2.4, Figures D2.4d and D2.4e). Additionally, Cluster 4.3 had greater average left occipital cortex volumes ($p<.05$) and reduced bilateral occipital and right temporal CSF volumes than Cluster 4.4 ($p<.05$; Table D2.4, Figure D2.4d).

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	Cluster 3 Mean	S.D.	Cluster 4 Mean	S.D.	ANCOVA P-value	Sig	F	Post Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=3,225)												
Executive Functioning 1	7.82E+00	2.48E+00	9.50E+00	3.53E+00	1.12E+01	2.91E+00	1.08E+01	3.12E+00	4.97E-04	***	6.14	4>1*,3>2*,3>1*
Executive Functioning 2	9.73E+00	2.20E+00	1.05E+01	3.05E+00	1.24E+01	2.34E+00	1.17E+01	2.97E+00	3.04E-04	***	6.51	3>1*,3>2**, 4>1*
Memory 1	8.55E+00	2.21E+00	9.80E+00	3.90E+00	1.16E+01	2.98E+00	1.13E+01	3.19E+00	1.44E-03	**	5.33	3>1*,3>2**
Memory 2	9.91E+00	2.26E+00	1.07E+01	3.57E+00	1.24E+01	2.62E+00	1.20E+01	2.86E+00	1.38E-03	**	4.37	3>1*,3>2**
Performance IQ 2	3.26E+01	1.03E+01	3.98E+01	1.26E+01	4.24E+01	1.18E+01	4.48E+01	1.17E+01	2.17E-03	**	5.03	4>1**,3>1*
Verbal IQ 1	3.63E+01	1.17E+01	4.70E+01	7.98E+00	4.55E+01	9.78E+00	4.62E+01	1.02E+01	7.24E-03	**	4.11	2>1**,3>1*, 4>1**
Verbal IQ 3	1.44E+01	4.06E+00	2.00E+01	3.11E+00	1.94E+01	4.28E+00	1.98E+01	4.38E+00	1.51E-04	***	4.11	4>1***,3>1** ,2>1***
Imaging Volumes (D.F.=3,162)												
Left Occipital Cortex	1.90E-02	2.30E-03	1.95E-02	2.50E-03	1.96E-02	2.10E-03	1.85E-02	2.30E-03	3.02E-02	*	5.05	3>4*
Right Parietal CSF	2.07E-02	2.80E-03	2.08E-02	5.10E-03	2.19E-02	4.90E-03	2.40E-02	4.80E-03	4.55E-03	**	4.51	4>2*
Left Parietal CSF	2.29E-02	4.56E-03	2.12E-02	4.52E-03	2.08E-02	4.98E-03	1.96E-02	3.61E-03	2.24E-02	*	3.28	N.S.
Right Occipital CSF	4.64E-03	1.40E-03	5.37E-03	1.50E-03	4.89E-03	1.20E-03	5.53E-03	1.40E-03	2.83E-02	*	3.1	4>3*
Left Occipital CSF	4.41E-03	1.30E-03	5.12E-03	1.20E-03	4.70E-03	1.10E-03	5.53E-03	1.50E-03	8.43E-04	***	5.82	4>3*
Right Temporal CSF	1.43E-02	2.70E-03	1.44E-02	2.50E-03	1.44E-02	2.20E-03	1.57E-02	2.90E-03	4.76E-03	**	4.48	4>3*
Cortisol Measures (D.F.=3,148)												
Baseline AM Cortisol Day 1 of DST	3.10E+00	4.10E-01	3.07E+00	7.30E-01	3.30E+00	5.10E-01	3.44E+00	4.60E-01	2.06E-02	*	3.35	4>2*

D.F., degrees of freedom
S.D., standard deviation
N.S., not significant
TK, Tukey-Kramer
*, p<.05; **, p<.01; ***, p<.001

Table D2.4. Biomarker Differences Across Clusters in the GWS Dataset for k=4

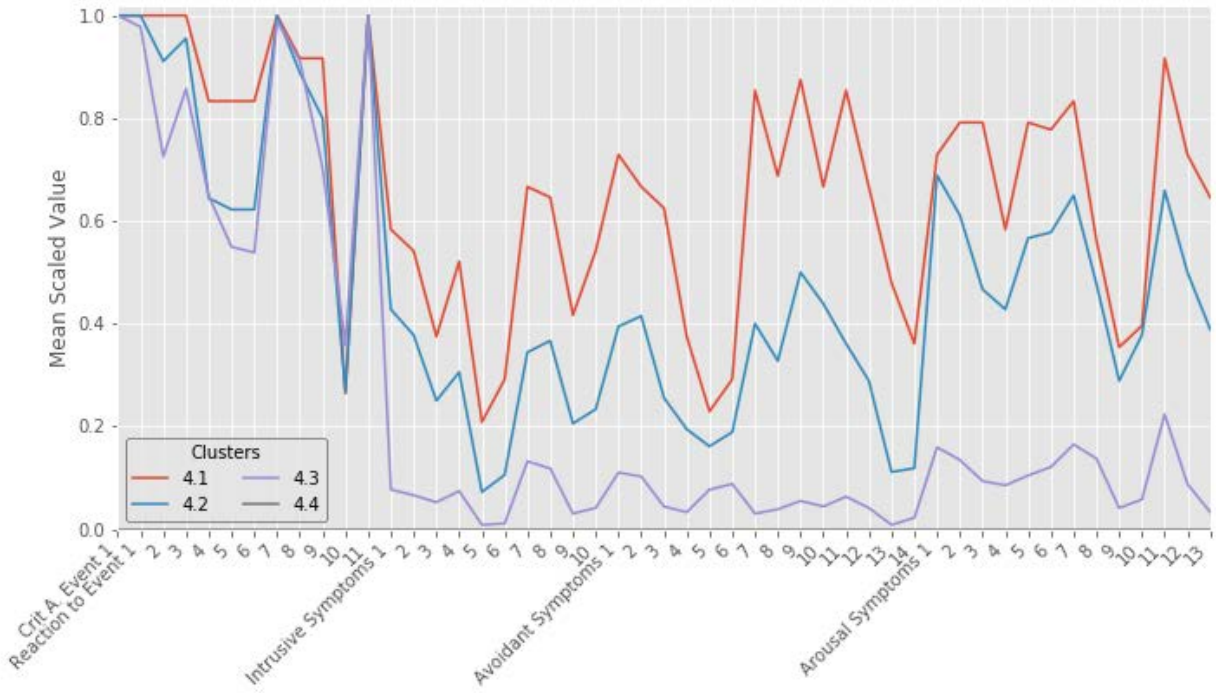

Figure D2.4a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=4)

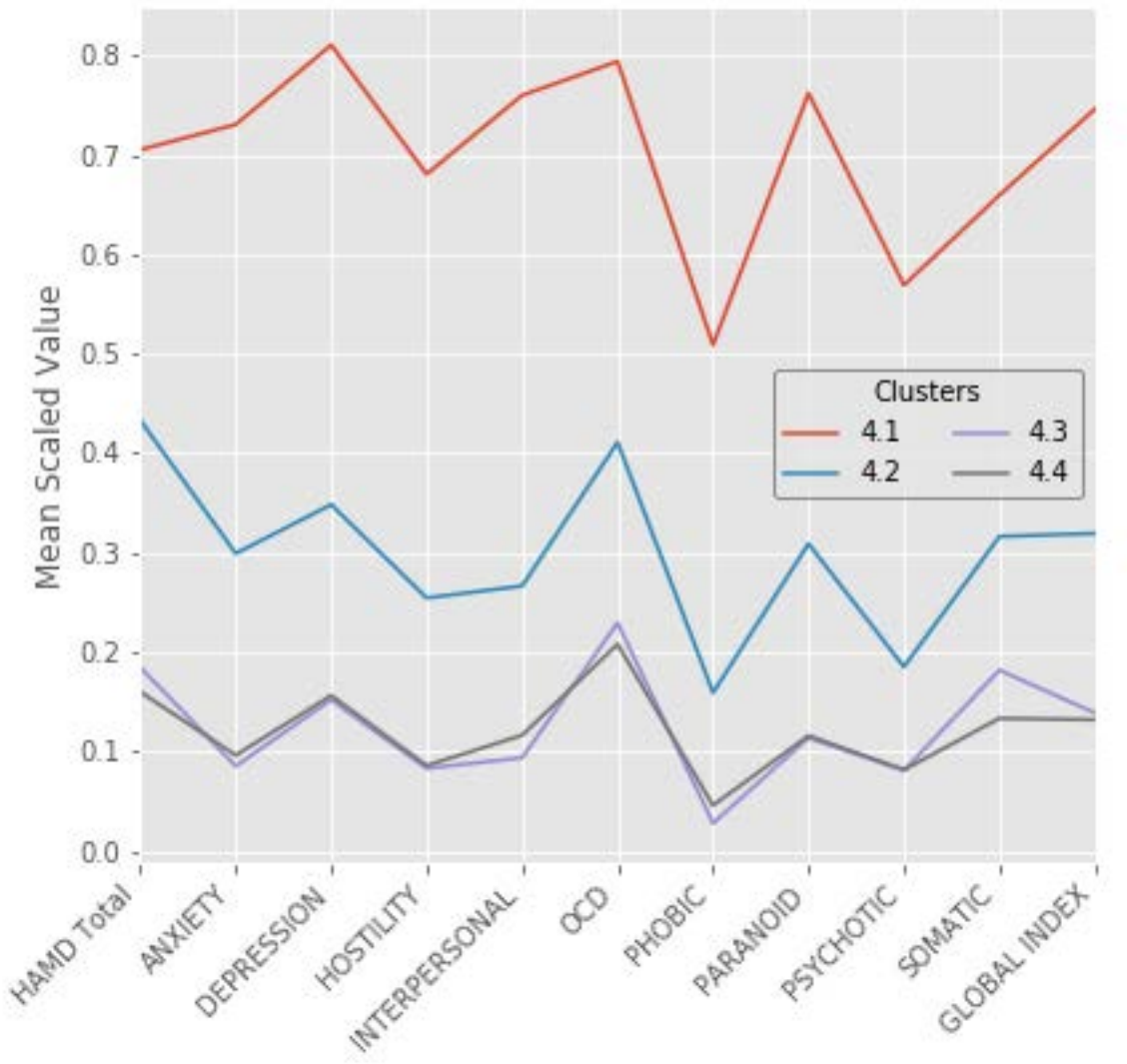

Figure D2.4b. Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 (k=4)

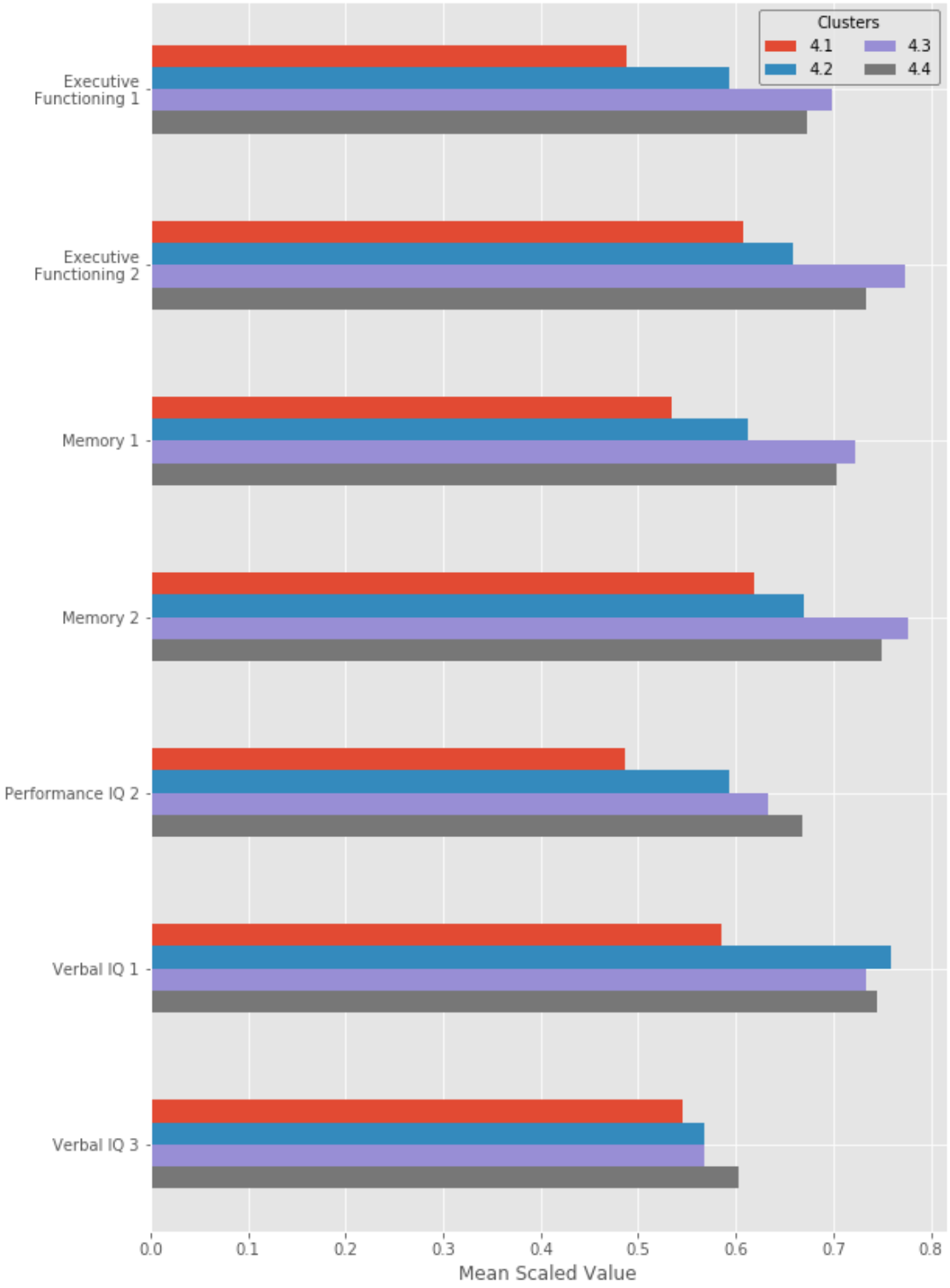

Figure D2.4c. Neuropsychiatric Markers with Significant Differences Across Clusters (k=4)

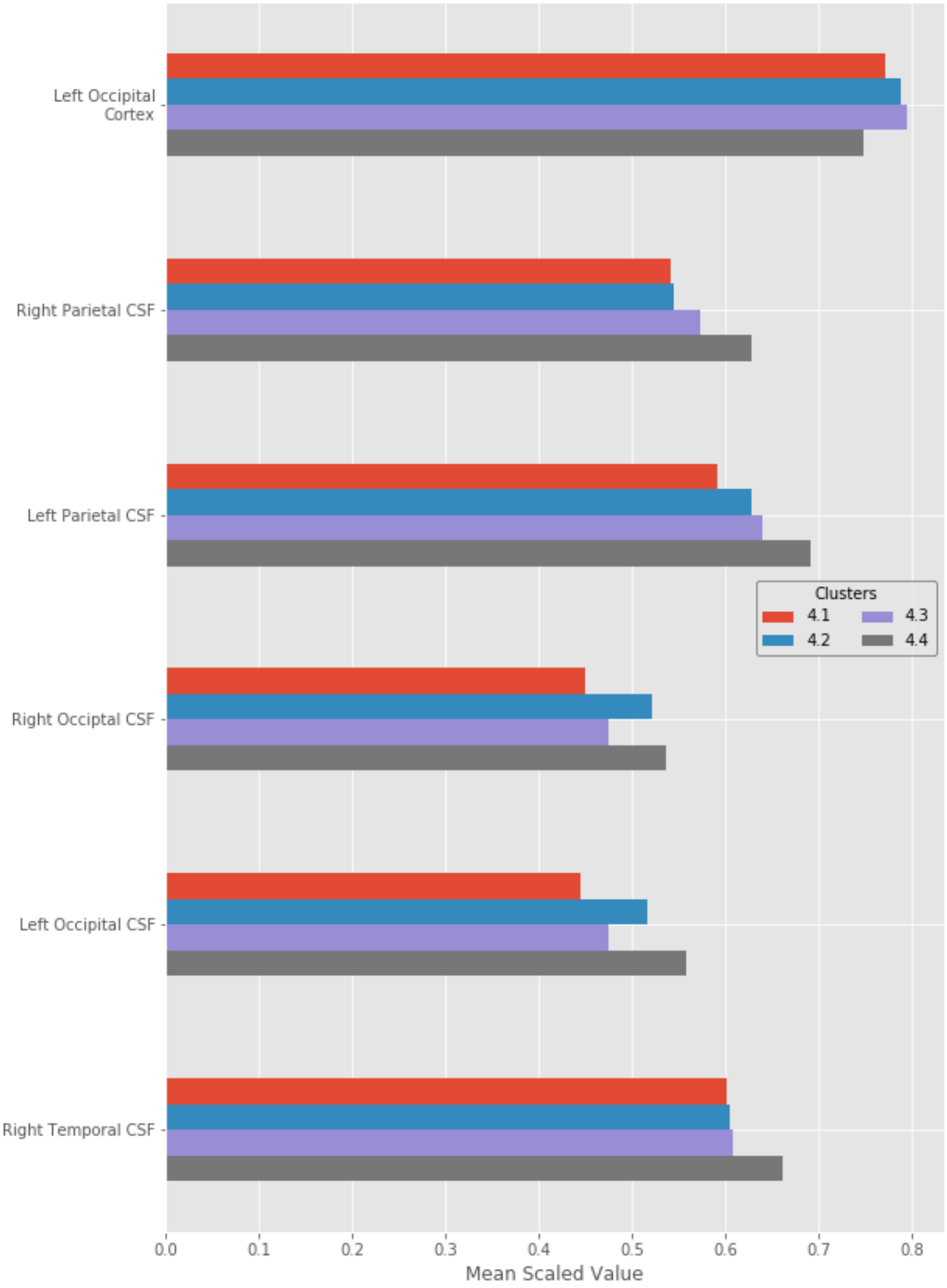

Figure D2.4d. Imaging Markers with Significant Differences Across Clusters (k=4)

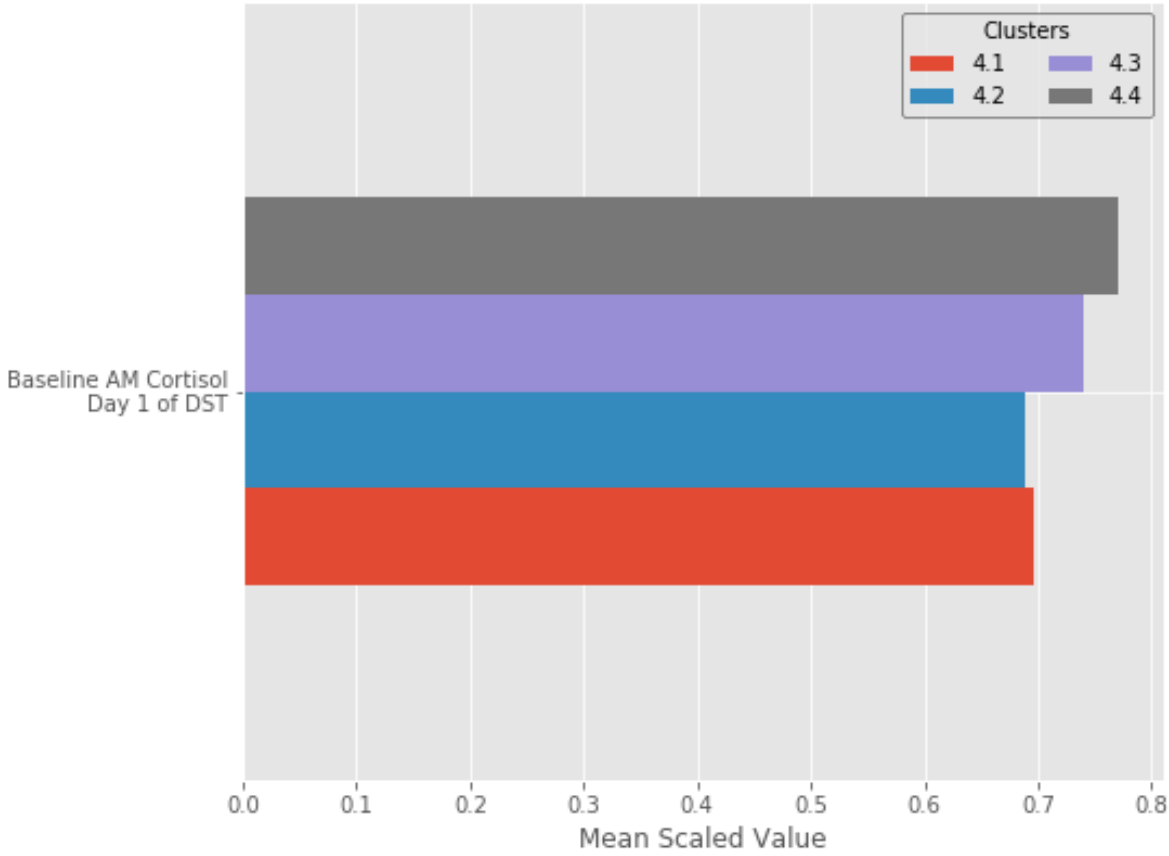

Figure D2.4e. Endocrine Markers with Significant Differences Across Clusters (k=4)

The five-cluster solution had the highest ARI (0.9), and all of the clusters were stable at least 87 percent of the time. As can be seen in Table D2.1, 30 percent (10/33) of individuals with PTSD composed the entire Cluster 5.1 (mean 3.2 DSM diagnoses). Cluster 5.2 comprised 52 percent (17/33) of individuals with PTSD (mean 2.1 DSM diagnoses); the remaining 17 percent of PTSD patients were in Cluster 5.3 (which had a mean 1.4 DSM diagnoses).

Cluster 5.1 had the highest CAPS symptom profile, followed by Clusters 5.2, 5.3, and 5.4, although there were several items within intrusive, numbing, and avoidant question subsets where individuals in Cluster 5.2 scored higher than those in Cluster 5.1 (Figure D2.5a). However, in the Ham-D and SCL-90 summary scores, symptom scores were uniformly higher in Cluster 5.1, followed by 5.2, and then 5.3 (Figure 6.5b). Individuals in Clusters 5.4 and 5.5 had lower scores across all of the summary domains than Clusters 5.1, 5.2, and 5.3 (Figure 6.5b). The incidence of alcohol dependence across all five clusters differed significantly ($\chi 2=14.2$, $p<.0001$; Table D2.1), with the percentage of individuals with alcohol dependence being the greatest in 5.2, followed by 5.1, 5.3, 5.4, and 5.5. A decreasing pattern as the cluster symptom profile decreased was also seen in the incidence of childhood trauma ($\chi 2=17.3$, $p<.01$; Table D2.1), which was highest in Cluster 5.1 and lowest in Cluster 5.5.

There were several differences between clusters in the neuropsychiatric domains. Individuals in Cluster 5.2 performed significantly worse than those in Clusters 5.4 and 5.5 in Executive Functioning and Memory ($p<0.05$ and $p<0.01$, respectively), while individuals in Cluster 5.1 performed significantly worse than individuals in Clusters 5.4 and 5.5 in Performance IQ and Verbal IQ ($p<0.05$; Table D2.5, Figure D2.5c). Cluster 5.2 had larger average right caudate white matter volumes than Clusters 5.4 and 5.3 ($p<0.01$) and larger average right lenticular white matter volumes than Clusters 5.3 and 5.5 ($p<0.05$; Table D2.5, Figure D2.5d).

There were also imaging differences between the two clusters that contained no patients with PTSD. There were no significant differences between Clusters 5.1, 5.3, and 5.2 in any of the measured imaging volumes (Table D2.5, Figure D2.5d). Finally, Cluster 5.3 had lower baseline cortisol than Cluster 5.4 ($p<0.05$; Table D2.5, Figure D2.5e).

Test	Cluster 1 Mean	S.D.	Cluster 5-9 Mean	S.D.	Cluster 5.2 Mean	S.D.	Cluster 4 Mean	S.D.	Cluster 5 Mean	S.D.	ANCOVA P-value	Sig	F	Post Hoc Cluster Difference TK
Neuropsychiatric (D.F.=4,225)														
Executive Functioning 1	8.33E+00	2.45E+00	8.71E+00	9.69E+00	1.02E+01	9.31E+00	1.19E+01	2.85E+00	1.08E+01	9.12E+00	1.19E-03	**	4.69	4>2**,5>2*
Executive Functioning 2	1.00E+01	2.00E+00	9.79E+00	9.41E+00	1.19E+01	2.70E+00	1.25E+01	2.14E+00	1.17E+01	2.97E+00	1.92E-04	***	6	4>2***,5>2*
Memory 1	9.00E+00	1.80E+00	8.75E+00	4.21E+00	1.05E+01	9.65E+00	1.17E+01	2.74E+00	1.19E+01	9.19E+00	5.93E-04	***	5.17	4>2**,5>2**
Memory 2	1.02E+01	2.05E+00	9.63E+00	9.99E+00	1.16E+01	9.09E+00	1.26E+01	2.99E+00	1.20E+01	2.86E+00	1.99E-04	***	5.97	4>2***,5>2**
Memory 4	2.09E+01	6.25E+00	2.08E+01	8.62E+00	2.54E+01	6.99E+00	2.49E+01	7.57E+00	2.58E+01	8.15E+00	9.99E-02	*	2.65	5>2*
Performance IQ 1	5.52E+01	1.34E+01	6.79E+01	1.29E+01	6.66E+01	1.54E+01	6.78E+01	1.70E+01	7.00E+01	1.91E+01	4.49E-02	*	2.48	5>1*
Performance IQ 2	9.20E+01	1.12E+01	9.76E+01	1.18E+01	4.04E+01	1.20E+01	4.90E+01	1.15E+01	4.48E+01	1.17E+01	1.67E-03	**	4.48	5>1*
Verbal IQ 1	9.53E+01	1.27E+01	4.66E+01	7.40E+00	4.44E+01	1.06E+01	4.63E+01	9.08E+00	4.62E+01	1.02E+01	1.67E-02	*	9.09	4>1*,5>1*4
Verbal IQ 3	1.49E+01	4.55E+00	1.99E+01	9.98E+00	1.92E+01	9.72E+00	1.98E+01	4.24E+00	1.98E+01	4.98E+00	2.50E-03	**	4.24	2>1*,3>1*,4>1**,5>1**
Imaging Volumes (D.F.=4,162)														
Right Occipital Cortex	1.92E-02	2.00E-03	1.88E-02	2.60E-03	1.86E-02	1.80E-03	1.94E-02	2.00E-03	1.85E-02	2.20E-03	2.91E-02	*	2.77	4>5*
Left Occipital Cortex	1.90E-02	2.90E-03	1.96E-02	2.80E-03	1.91E-02	1.90E-03	1.97E-02	2.20E-03	1.85E-02	2.90E-03	4.00E-02	*	2.57	4>5*
Right Caudate White Matter	7.48E-04	2.04E-04	1.07E-03	6.08E-04	6.72E-04	1.80E-04	7.14E-04	2.22E-04	7.49E-04	2.57E-04	2.46E-03	**	4.91	2>3**,2>4**,2>5**
Right Lenticular White Matter	2.47E-03	4.19E-04	2.59E-03	7.46E-04	2.11E-03	9.86E-04	2.27E-03	4.70E-04	2.21E-03	4.22E-04	1.92E-02	*	2.57	2>3*,2>5*
Left Parietal CSF	1.96E-02	9.61E-03	2.09E-02	5.15E-03	2.20E-02	5.15E-03	2.09E-02	4.93E-03	2.25E-02	4.56E-03	2.52E-02	*	2.86	NS.
Right Occipital CSF	4.64E-03	1.40E-03	5.24E-03	1.90E-03	5.55E-03	1.80E-03	4.81E-03	1.10E-03	5.55E-03	1.40E-03	2.16E-02	*	2.96	5>4*
Left Occipital CSF	4.41E-03	1.90E-03	4.92E-03	1.20E-03	5.40E-03	1.90E-03	4.59E-03	1.10E-03	5.55E-03	1.50E-03	9.09E-04	***	5.56	5>4**
Right Temporal CSF	1.49E-02	2.70E-03	1.49E-02	2.90E-03	1.48E-02	2.60E-03	1.49E-02	2.00E-03	1.57E-02	2.90E-03	8.04E-03	**	3.57	5>4*
Left Temporal CSF	1.34E-02	2.56E-03	1.28E-02	2.47E-03	1.96E-02	2.55E-03	1.33E-02	3.15E-03	1.44E-02	3.07E-03	8.04E-03	**	3.57	4>2*
Subcortical CSF	9.45E-04	1.22E-04	1.00E-03	9.04E-04	1.04E-03	2.72E-04	9.98E-04	1.95E-04	1.08E-03	2.93E-04	1.80E-02	*	3.08	4>2**
Cerebellar CSF	1.90E-02	5.40E-03	1.93E-02	4.79E-03	2.19E-02	4.60E-03	1.84E-02	3.60E-03	2.05E-02	4.58E-03	3.50E-02	*	2.65	NS.
Cortisol Measures (D.F.=4,147)														
Baseline AM Cortisol Day 1 of DST	9.05E+00	4.50E-01	9.28E+00	5.40E-01	2.94E+00	6.90E-01	9.38E+00	5.30E-01	9.44E+00	5.00E-01	4.97E-03	**	3.89	4>3*,5>3**

D.F., degrees of freedom
S.D., standard deviation
N.S., not significant
TK, Tukey-Kramer
*, p< .05; **, p< .01; ***, p< .001

Table D2.5. Biomarker Differences Across Clusters in the GS Dataset for k=5

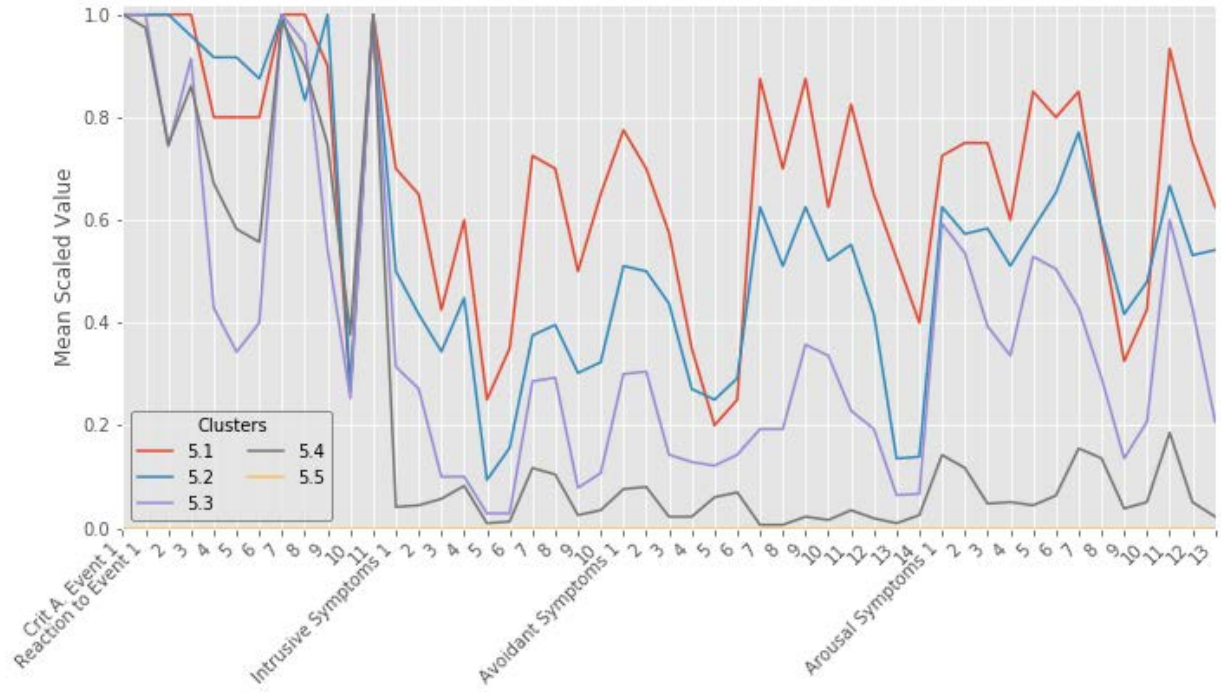

Figure D2.5a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=5)

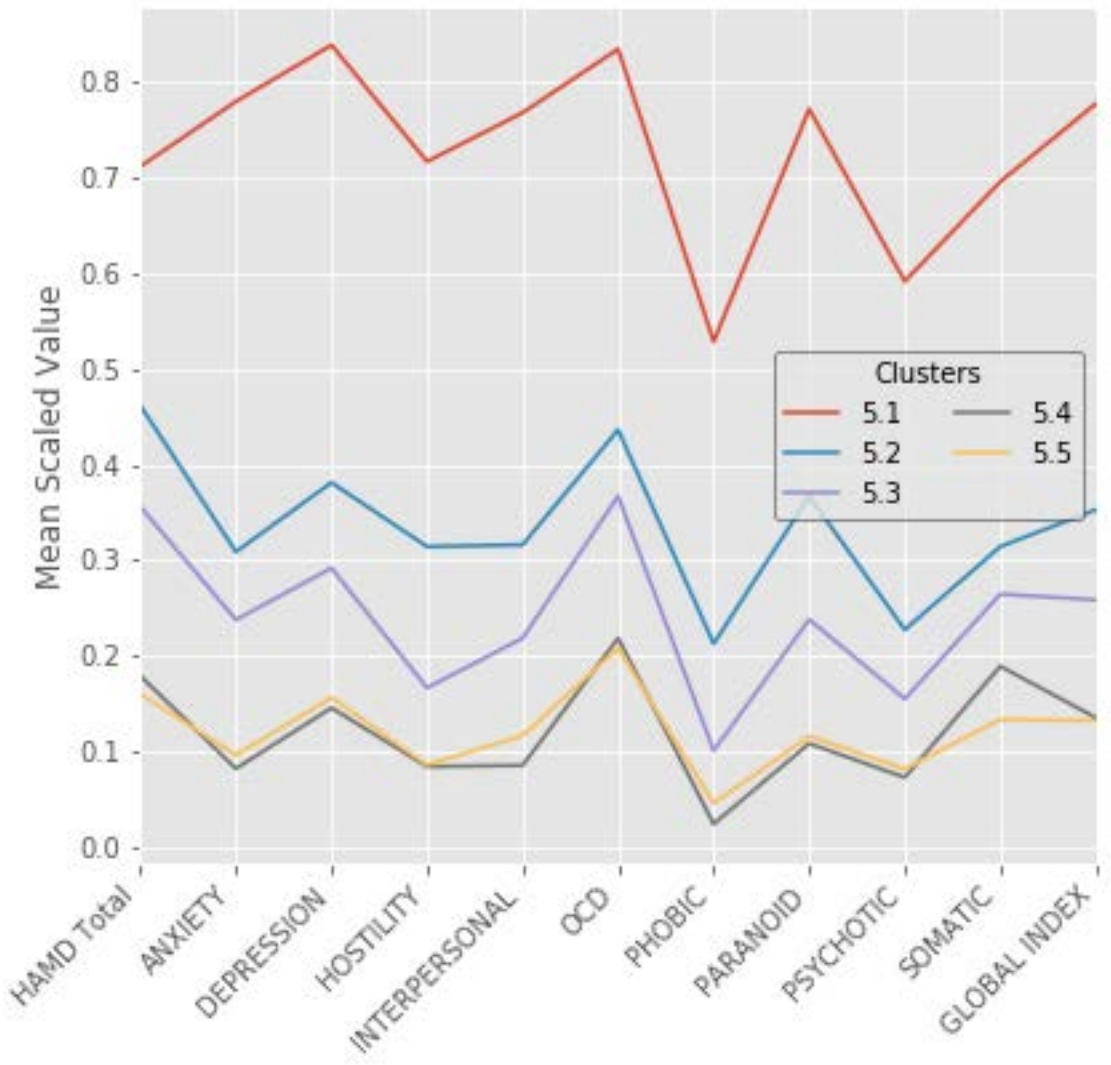

Figure D2.5b. Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 (k=5)

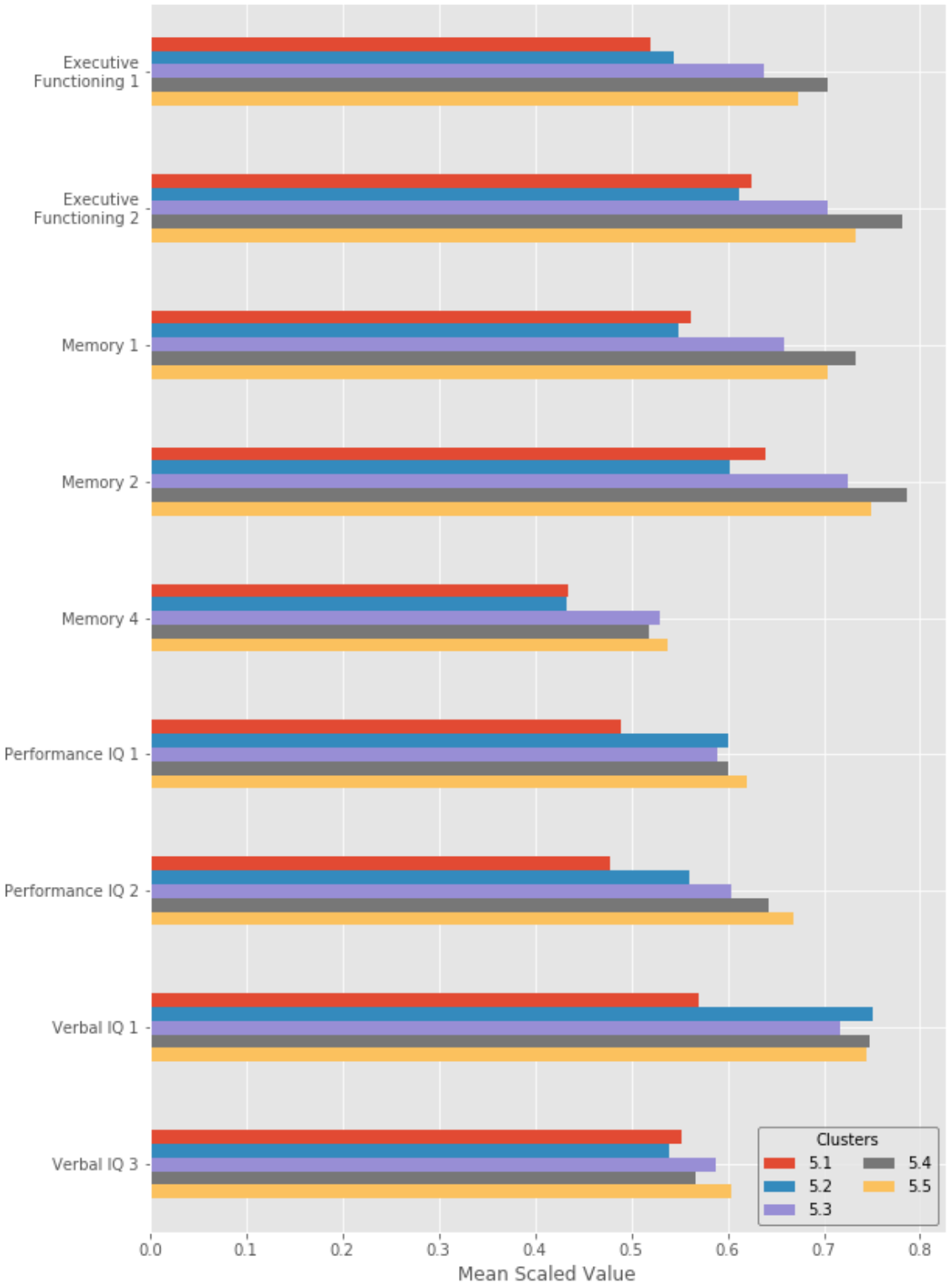

Figure D2.5c. Neuropsychiatric Markers with Significant Differences Across Clusters (k=5)

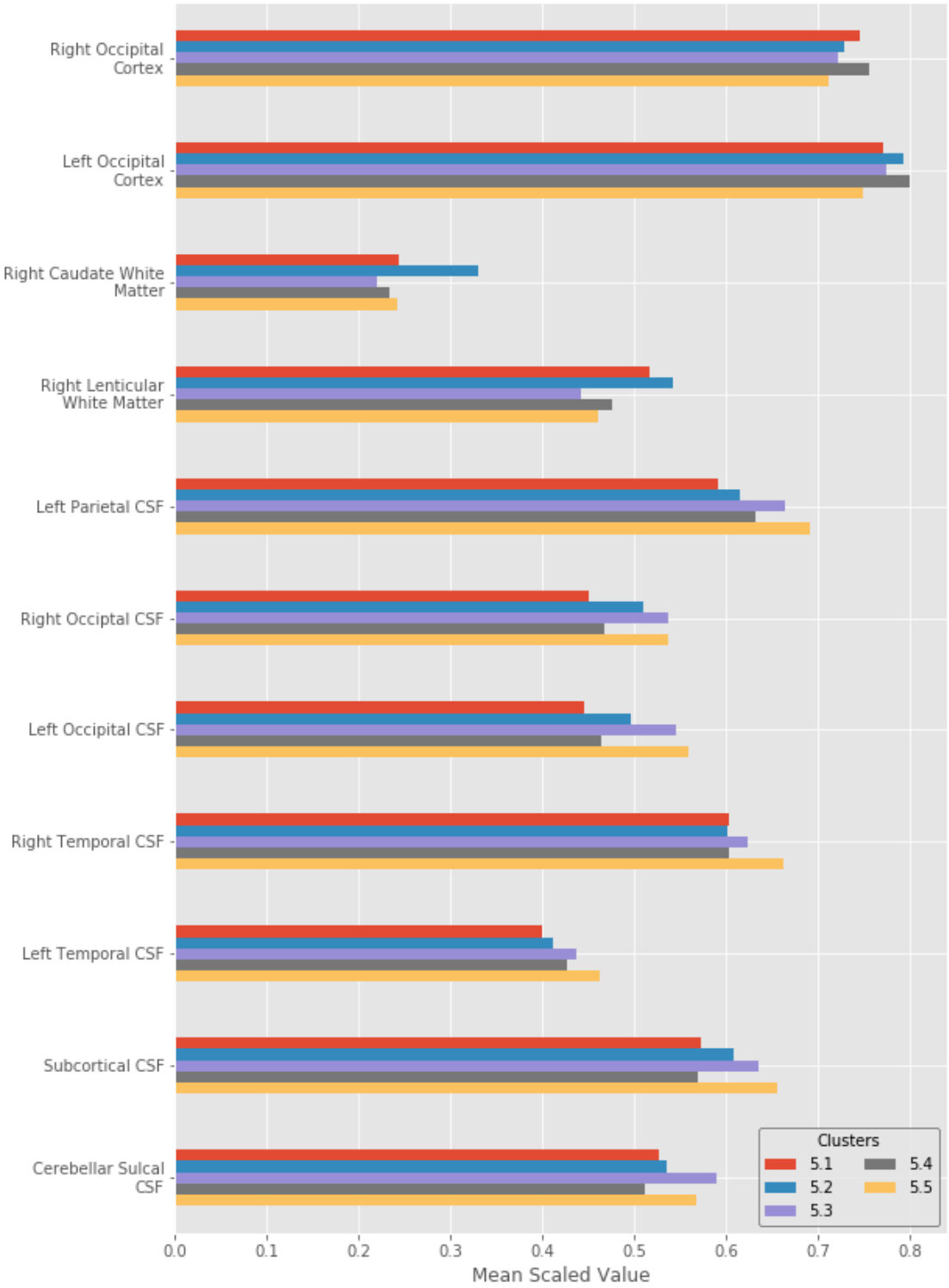

Figure D2.5d. Imaging Markers with Significant Differences Across Clusters (k=5)

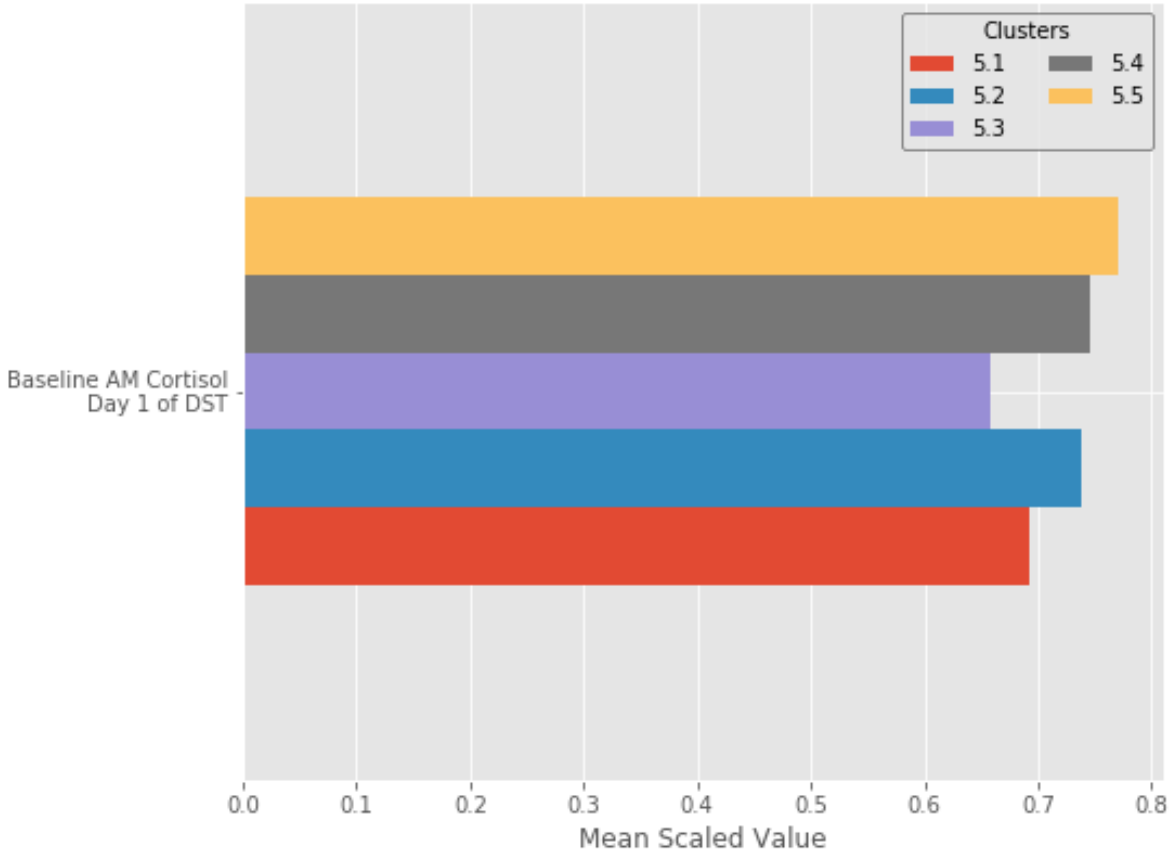

Figure D2.5e. Endocrine Markers with Significant Differences Across Clusters (k=5)

APPENDIX D3: Cluster Analyses of MIRECC Summary (MS) Dataset for k=2, k=3, k=4, and k=5 Solutions, with Symptom Profiles and Biomarker Differences

This appendix contains the symptom profiles derived using the MIRECC Summary (MS) dataset to cluster individuals, using the full CAPS and summary scores from the Ham-D and the SCL-90. The MIRECC dataset was clustered with the same summary set of 61 psychiatric symptoms as described in the GS from the methods section and Appendix D2. For the MIRECC Summary (MS) dataset, there were 84 male patients with full psychiatric symptom data available for use in the cluster analysis, and 50 with PTSD diagnoses. The cluster descriptions for the k=2, k=3, k=4, and k=5 analyses are shown in Table D3.1. Biomarker differences across clusters for all four solutions are also presented adjacent to the symptom profiles. The Castleman-Hasslebach (CH) and wss graph is located in Figure D3.1.

Table D3.1 Details of Cluster Solutions for k=2, k=3, k=4, and k=5

Dataset (# of Clusters) ARI Bootstrap	Cluster (# of Individuals in Cluster)	PTSD	MDD	Other DSM Avg(Range)	Total DSM Avg(Range)	ALC Dependence	$\chi 2$	Childhood Trauma	$\chi 2$
MS (2)	Cluster2.1 (49)	49	16	1.9(0-8)	3.7(1-11)	25	0.23	18	5.2*
ARI: 0.86	Cluster2.2 (35)	1	2	1.3(0-9)	1.8(0-9)	16		5	
Boot: 0,0									
MS (3)	Cluster3.1 (16)	16	8	1.2(0-4)	3.1(1-6)	16	23*****	5	5.5
ARI: 0.60	Cluster3.2 (33)	33	8	2.2(0-8)	4.0(1-11)	9		13	
Boot: 7,010	Cluster3.3 (35)	1	2	1.3(0-9)	1.8(0-9)	16		5	
MS (4)	Cluster4.1 (13)	13	7	1.1(0-3)	3.0(0-9)	5	3.87	4	5.5
ARI: 0.80	Cluster4.2 (15)	15	5	1.9(0-7)	3.6(1-6)	6		6	
Boot: 60,4,0,16	Cluster4.3 (21)	21	4	2.3(0-8)	4.2(1-11)	14		8	
	Cluster4.4 (35)	1	2	1.3(0-9)	1.8(0-9)	16		5	
	Cluster5.1 (13)	13	7	1.1(0-3)	1.8(1-6)	5		4	
MS (5)	Cluster5.2 (15)	15	5	1.9(0-7)	3.6(1-9)	6	5	6	5.6
ARI: 0.80	Cluster5.3 (21)	21	4	2.3(0-8)	4.2(1-11)	14		8	
Boot: 12,11,36,0,2	Cluster5.4 (34)	1	2	1.1(0-9)	1.6(0-9)	15		5	
	Cluster5.5 (1)	0	0	6(6-6)	7(7-7)	1		0	
	Cluster2.1 (49)	49	16	1.9(0-8)	3.7(1-11)	25		18	

MS: MIRECC Summary
ARI: Adjusted Rand Index between the k-means cluster and hierarchical cluster solutions
BS: Bootstrap shows the number of times each cluster dissolved during 100 bootstraps of the cluster analyses.
*, $p<.05$; *****, $p<.00001$

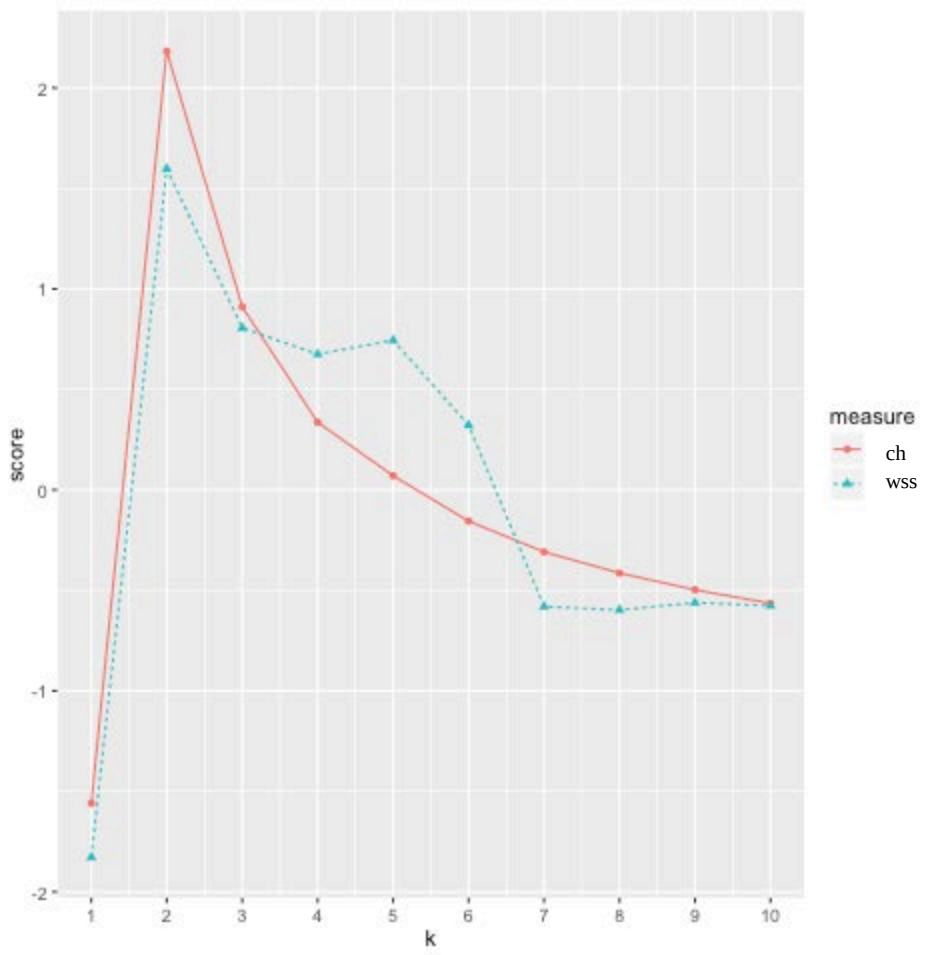

Figure D3.1. Details of Cluster Solutions for k=2, k=3, k=4, and k=5

The two-cluster solution was very robust, as shown by the ARI and the bootstrap criteria. Cluster 2.1 contained 49 of the 50 individuals with PTSD and the majority of individuals with MDD. As shown in Figures D3.2a and D3.2b, Cluster 2.1 had higher levels of all items in the CAPS, Ham-D, and the SCL-90 summary scores. Patients with alcohol dependence were less clearly separated in these clusters, but individuals in Cluster 2.1 had a significantly greater incidence of childhood trauma than those in Cluster 2.2 ($\chi 2=5.3$, $p<.05$; Table D2.1).

Table D3.2 and Figure D3.2c show that individuals in Cluster 2.1 exhibited several significant neuropsychiatric impairments, in the Learning ($p<.01$, $p<.001$), Performance IQ ($p<.05$), and Verbal IQ ($p<.001$) domains. Individuals in this cluster also displayed smaller average volumes in the left and right frontal cortices, as well as the right temporal cortex, as shown in Figure D3.2d ($p<.05$ for all). There were no differences in cortisol measures.

Table D3.2. Biomarker Differences Across Clusters in MS Dataset for k=2

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	ANCOVA P-value	Sig	F	Post-Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=1,80)								
Learning 1	5.73E+00	1.55E+00	7.03E+00	1.89E+00	7.10E-04	***	12.41	1<2 ***
Learning 2	1.05E+01	2.61E+00	1.23E+01	2.21E+00	1.84E-03	**	10.39	1<2 **
Learning 3	4.38E+01	1.03E+01	5.17E+01	8.85E+00	2.90E-04	***	14.37	1<2 ***
Learning 4	5.69E+00	1.50E+00	6.49E+00	1.92E+00	3.89E-02	*	0.11	1<2 *
Performance IQ 1	6.03E+01	1.28E+01	6.73E+01	1.47E+01	1.56E-02	*	6.11	1<2 *
Performance IQ 2	3.69E+01	1.07E+01	4.27E+01	1.25E+01	2.10E-02	*	5.56	1<2 *
Verbal IQ 3	1.51E+01	3.48E+00	1.81E+01	3.92E+00	3.52E-04	***	13.94	1<2***
Imaging Volumes (D.F.=1,66)								
Right Frontal Cortex	7.89E-02	8.60E-03	8.27E-02	6.30E-03	2.60E-02	*	5.18	1<2 *
Left Frontal Cortex	7.92E-02	8.20E-03	8.33E-02	6.70E-03	2.04E-02	*	5.64	1<2 *
Right Temporal Cortex	4.67E-02	5.50E-03	4.93E-02	4.30E-03	2.74E-02	*	5.09	1<2 *

D.F.: degrees of freedom
S.D.: standard deviation
TK: Tukey-Kramer
*, p<.05; **, p<.01; ***, p<.001

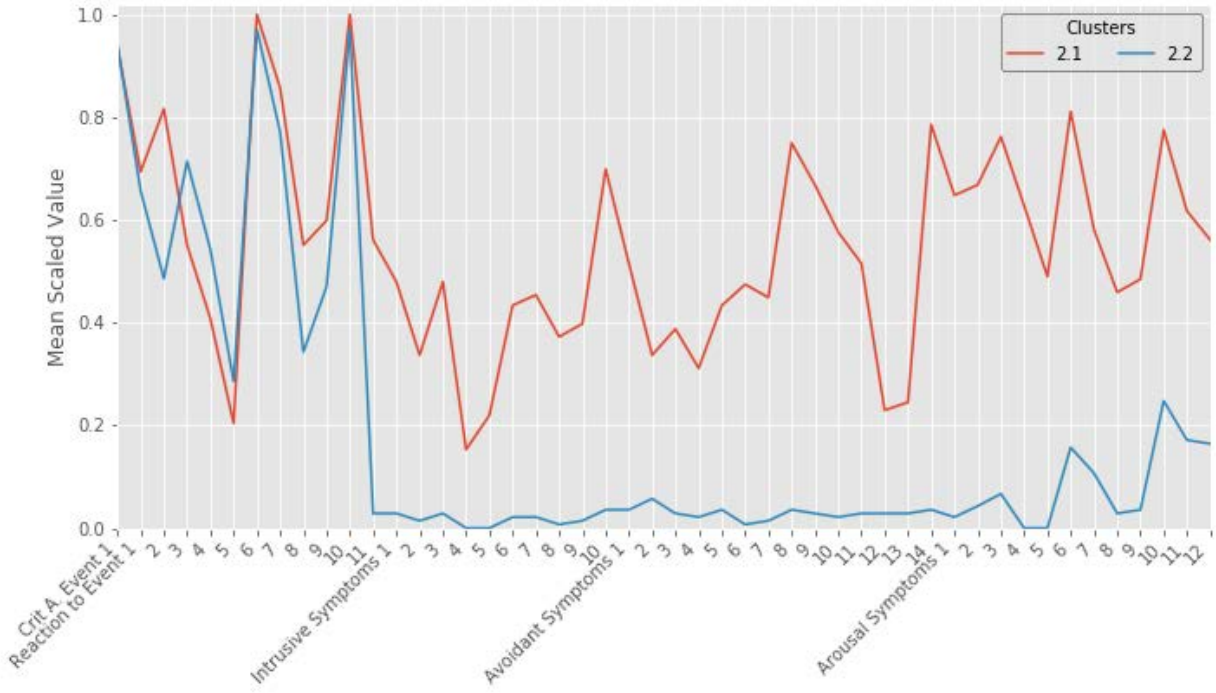

Figure D3.2a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=2)

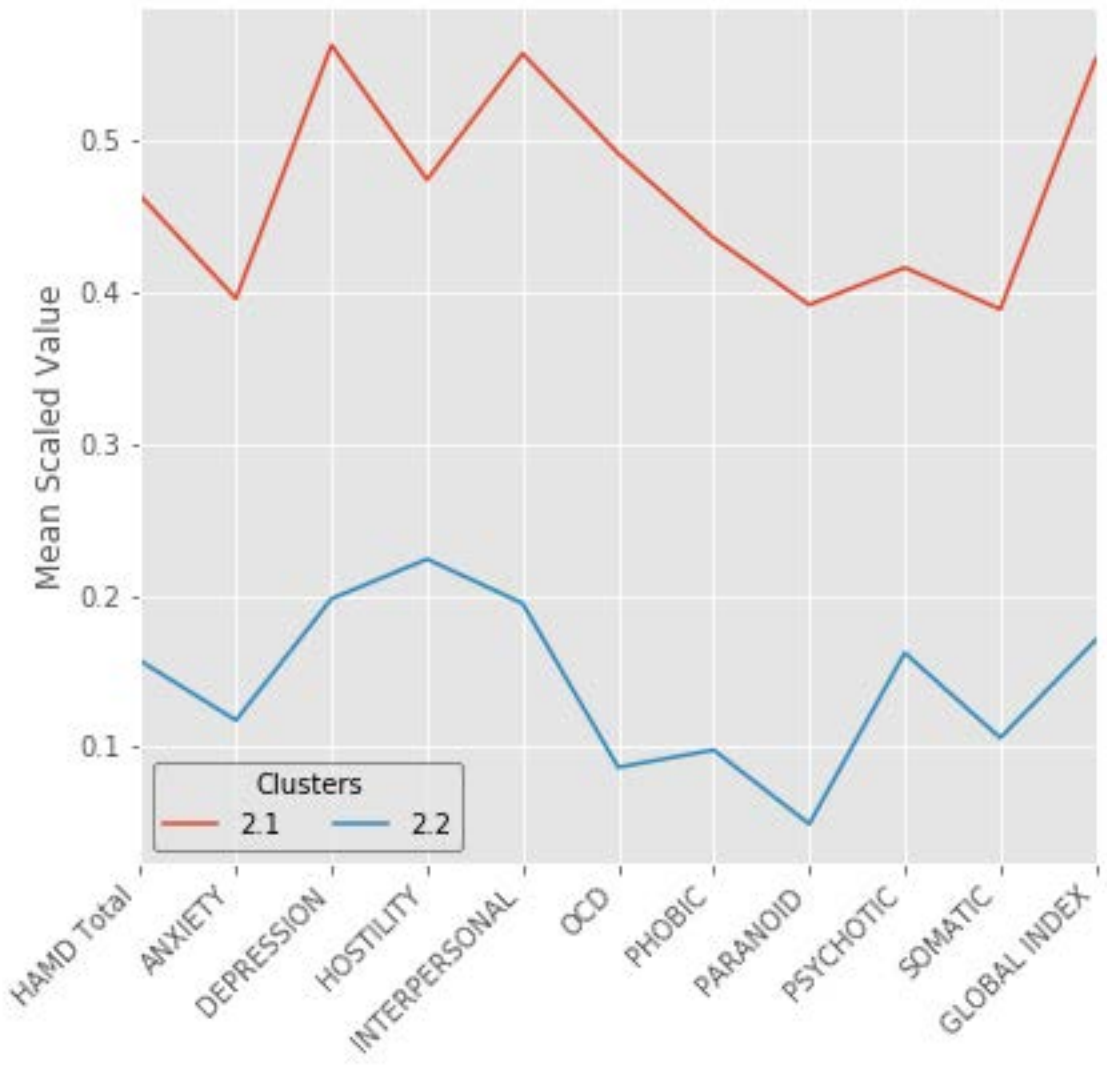

Figure D3.2b. Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 (k=2)

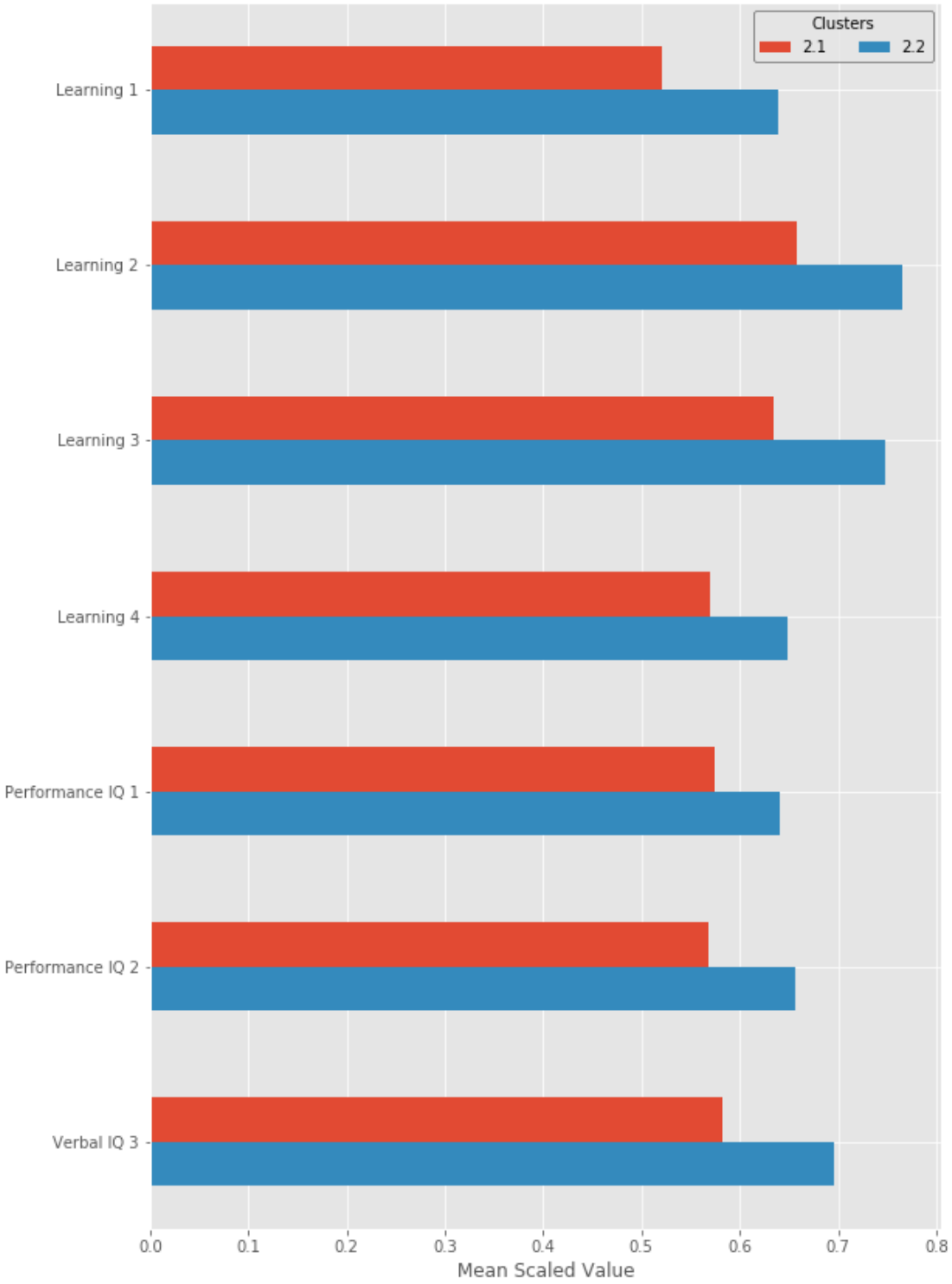

Figure D3.2c. Neuropsychiatric Markers with Significant Differences Across Clusters (k=2)

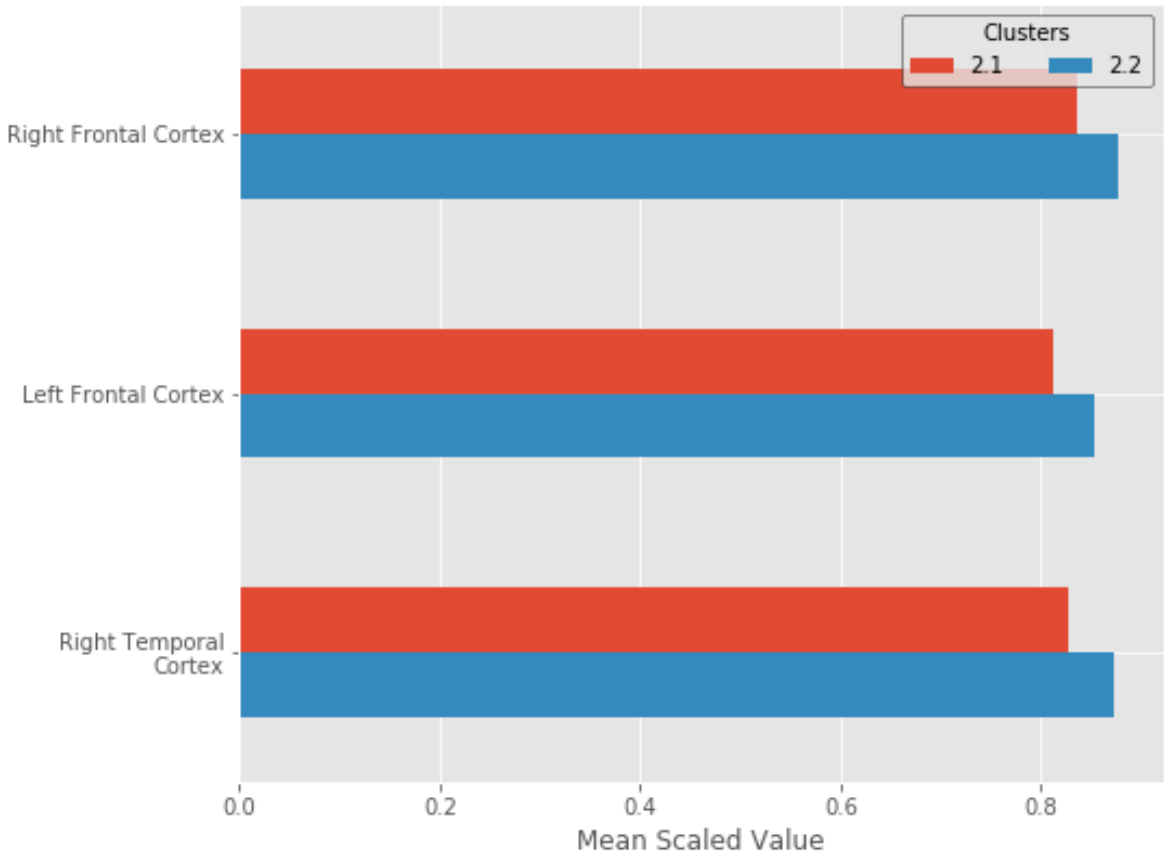

Figure D3.2d. Imaging Markers with Significant Differences Across Clusters (k=2)

The three-cluster solution had a lower ARI of 0.6, but the bootstrap analysis showed that the clusters were all stable greater than 93 percent of the time. In this solution, Cluster 3.3 emerged identical to Cluster 2.2 (Table D3.1), while Clusters 3.1 and 3.2 resulted from a split of Cluster 2.1. Of the three clusters, Cluster 3.3 had the fewest psychiatric symptoms, followed by Cluster 3.2 and Cluster 3.1 (Figures D3.3a and D3.3b). The incidence of alcohol dependence was 100 percent in Cluster 3.1, and differed very significantly from the incidence in both Clusters 3.2 and 3.3 ($\chi 2=23$, $p<.00001$; Table D3.1). There was no difference in the incidence of childhood trauma across clusters.

Individuals in Cluster 3.3 performed significantly better on neuropsychiatric tests in the domains of Learning, Performance IQ, and Verbal IQ ($p<.05$, $p<.01$) than individuals in Clusters 3.2 and 3.1 (Table D3.3, Figure D3.3c). Finally, Cluster 3.1 exhibited significant hyposuppression of cortisol on Day 2 of the DST compared with Cluster 3.3 ($p<.05$; Table D3.3, Figure D3.3c).

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	Cluster 3 Mean	S.D.	ANCOVA P-value	Sig	F	Post Hoc Cluster Difference
Neuropsychiatric (D.F.=2,79)										
Learning 1	5.88E+00	1.36E+00	5.67E+00	1.65E+00	7.03E+00	1.89E+00	3.07E-03	**	6.23	3>2**,3>1*
Learning 2	1.01E+01	3.50E+00	1.07E+01	2.08E+00	1.23E+01	2.21E+00	5.82E-03	**	5.49	3>2*,3>1**
Learning 3	4.18E+01	1.30E+01	4.47E+01	8.74E+00	5.17E+01	8.85E+00	8.46E-04	***	7.75	3>2*,3>1**
Performance IQ 1	5.53E+01	9.92E+00	6.27E+01	1.34E+01	6.73E+01	1.47E+01	9.25E-03	**	4.97	3>1**
Verbal IQ 3	1.53E+01	4.19E+00	1.51E+01	3.15E+00	1.81E+01	3.92E+00	1.73E-03	**	6.89	3>1*
Imaging Measures (D.F.=2,65)										
Left Frontal Cortex	8.12E-02	7.41E-03	7.85E-02	8.46E-03	8.33E-02	6.65E-03	3.77E-02	*	3.45	N.S.
Cortisol Measures (D.F.=2,65)										
Percent Suppression Baseline Cortisol in DST(D1-D2/D1)	1.53E+01	4.19	1.51E+01	3.152	1.81E+01	3.92	1.73E-03	**	6.89	3>1*

D.F., degrees of freedom
S.D., standard deviation
N.S., not significant
TK, Tukey-Kramer
*, p<.05; **, p<.01; ***, p<.001

Table D3.3. Biomarker Differences Across Clusters in the MS Dataset for k=3

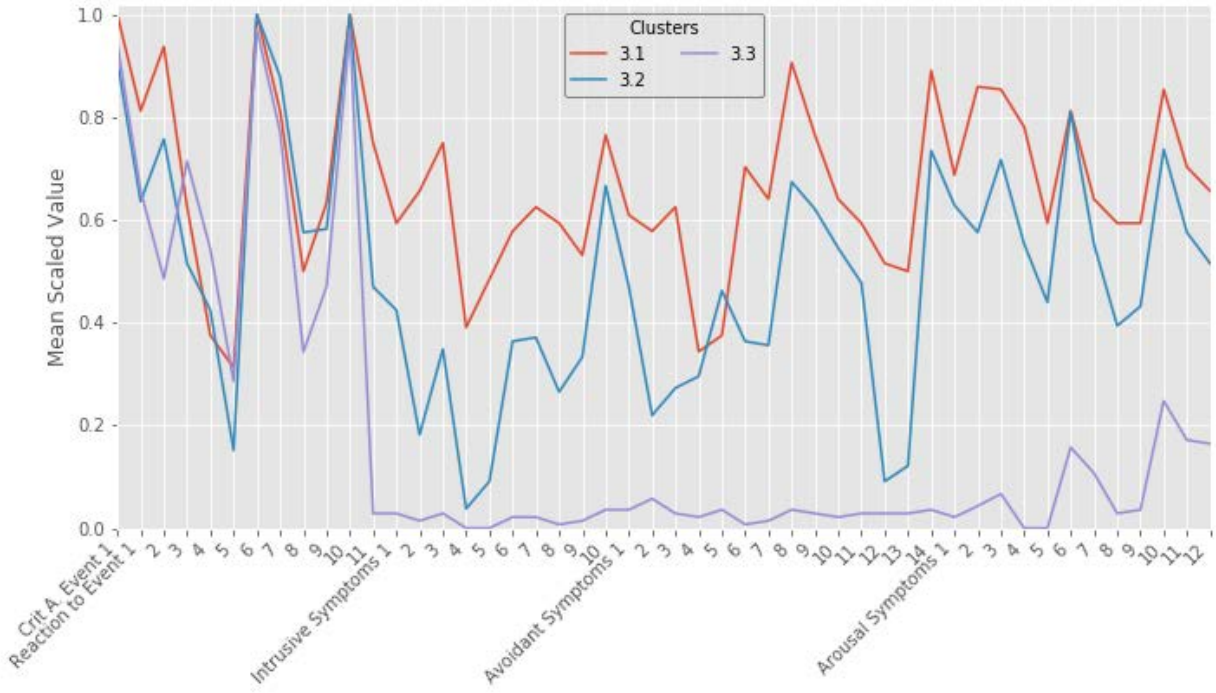

Figure D3.3a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=3)

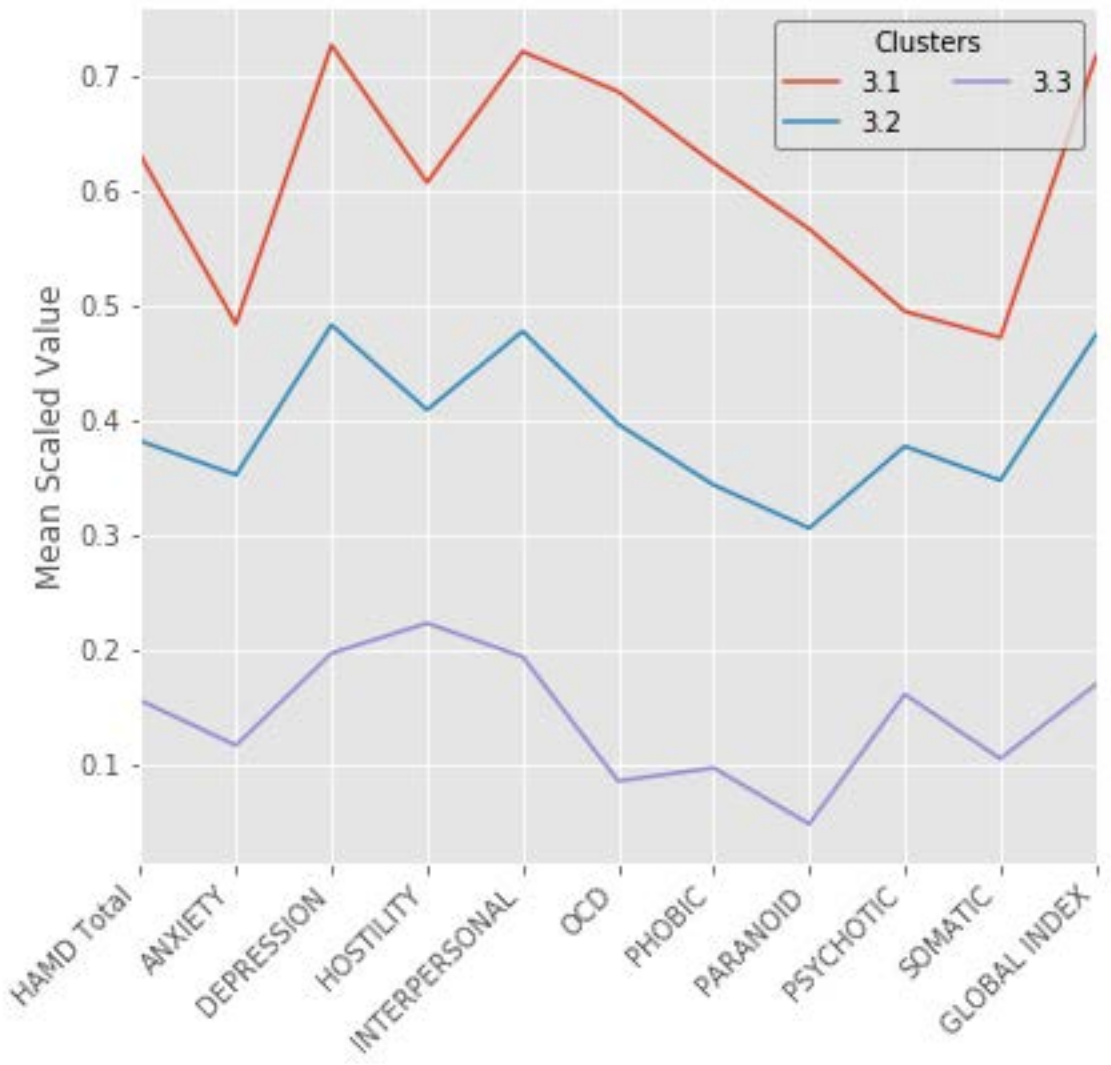

Figure D3.3b. Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 (k=3)

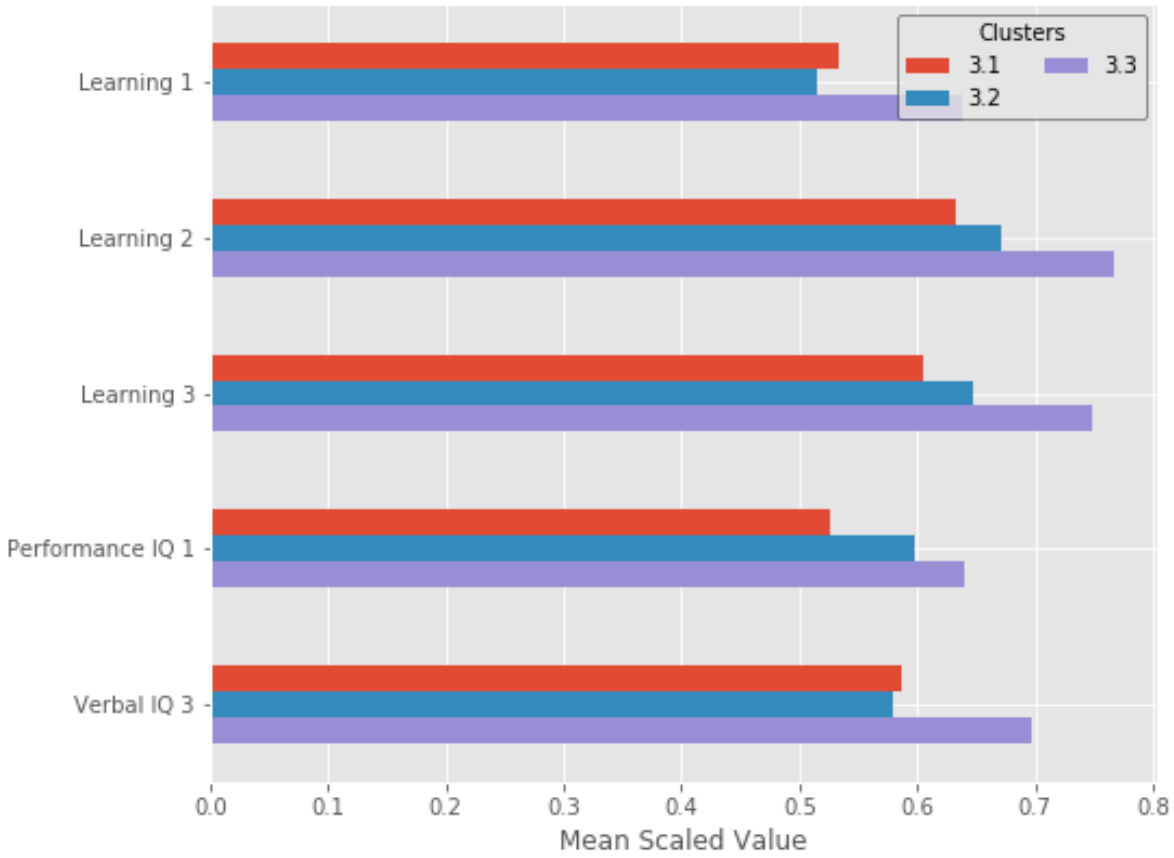

Figure D3.3c. Neuropsychiatric Markers with Significant Differences Across Clusters (k=3)

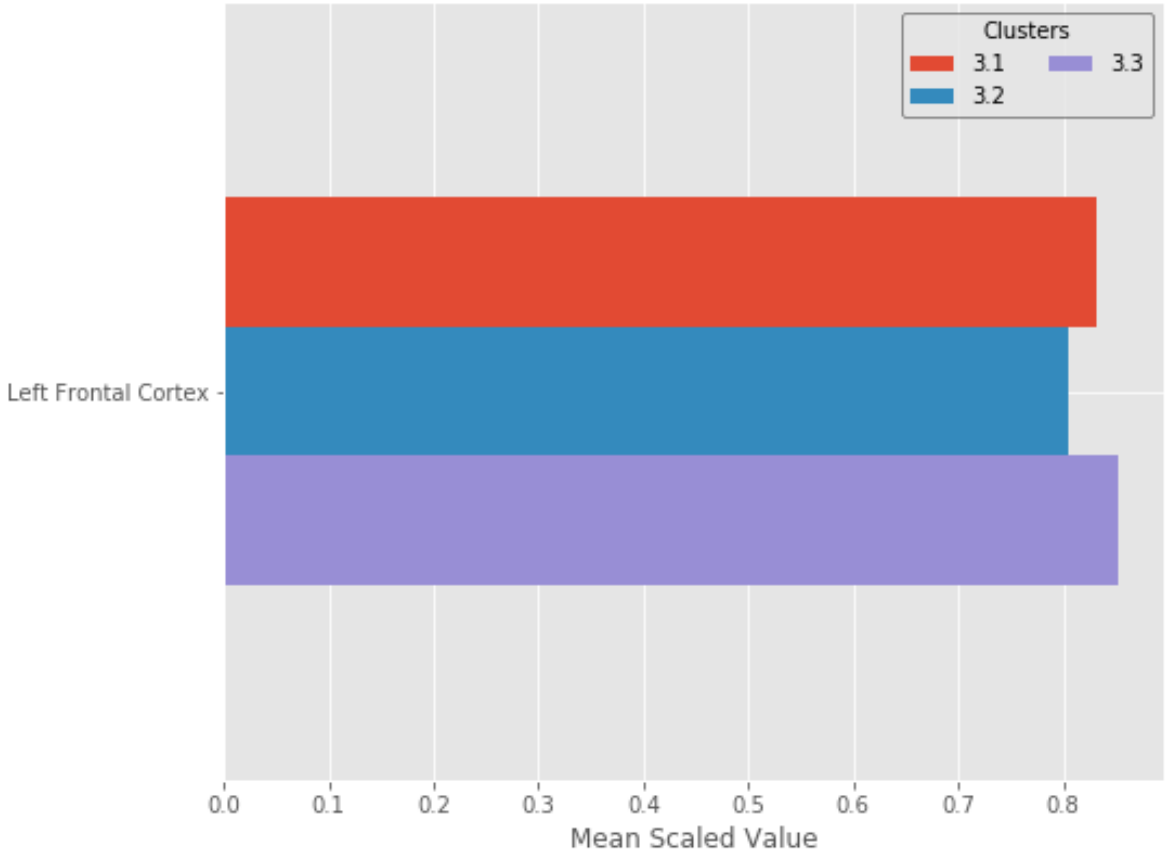

Figure D3.3d. Endocrine Markers with Significant Differences Across Clusters (k=3)

Although the ARIs were high for the four- and five-cluster solutions, the bootstrap measure fell below the 80 percent stability mark. As shown in Table D3.1, in both of these analyses, the individuals with PTSD clustered into three stable clusters that were identical between the four- and five-cluster solutions (i.e., Clusters 4.1 and 5.1 were identical, Clusters 4.2 and 5.2 were identical, and Clusters 4.3 and 5.3 were identical).

In the four-cluster solution the mean values across symptom profiles were inversely proportional to Cluster number, with Cluster 4.1 having the highest symptom profile, and 4.4 having the lowest (Figures D3.4a and D3.4b). The four-cluster solution also showed that the symptom profiles were inversely proportional to performance on neuropsychiatric tests (Figure D3.4c). Cluster 4.1, which had the highest symptom profile, also had the least suppression of cortisol, and individuals in this cluster had significantly lower Day 1 cortisol than individuals in Clusters 4.2 and 4.4 ($p<.01$, $p<.05$; Table D3.4, Figure D3.4d). As all individuals in Clusters 4.2 and 4.1 had PTSD diagnoses, these results demonstrate delineation in cortisol response within individuals with a DSM-IV diagnosis of PTSD. This difference also was present in the five-cluster solution, along with the differences between groups where the neuropsychiatric domains were inversely proportional to psychiatric symptom levels ($p<.05$, $p<.01$; Table D3.4, Figures D3.5a–d). The only difference between the k=4 and k=5 solutions was that one individual dropped out of Cluster 5.4 and formed Cluster 5.5 with only one member. Given the small size of Cluster 5.5, the five-cluster solution has little effect on the findings because there can be no statistical significance in a group with an N=1. Thus, the k=4 and k=5 solutions are essentially the same for the purpose of these analyses.

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	Cluster 3 Mean	S.D.	Cluster 4 Mean	S.D.	ANCOVA P-value	Sig	F	Post Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.= 3,78)												
Learning 1	5.92E+00	1.38E+00	6.07E+00	1.44E+00	5.38E+00	1.72E+00	7.03E+00	1.89E+00	4.62E-03	**	4.67	3<4**
Learning 2	1.06E+01	3.28E+00	1.09E+01	2.67E+00	1.02E+01	2.17E+00	1.23E+01	2.21E+00	1.71E-02	*	5.6	3<4*
Learning 3	4.23E+01	1.28E+01	4.67E+01	9.06E+00	4.25E+01	9.39E+00	5.17E+01	8.85E+00	1.66E-03	**	5.55	3<4**,1<4 *
Verbal IQ 3	1.60E+01	3.76E+00	1.45E+01	3.54E+00	1.50E+01	3.32E+00	1.81E+01	3.92E+00	3.33E-03	**	4.96	3<4*, 1<4*
Cortisol Measures (D.F.=3,64)												
Baseline AM Cortisol Day 1 of DST	2.89E+00	4.70E-01	3.75E+00	3.30E-01	3.32E+00	7.30E-01	3.37E+00	7.00E-01	5.28E-03	**	4.62	2>1**, 4>1*
Percent Suppression Baseline Cortisol in DST(D1-D2/D1)	4.41E+00	7.51E+00	1.62E+01	1.34E+01	1.54E+01	1.38E+01	1.50E+01	1.11E+01	1.79E-02	*	3.1	3>1*,4>1**

D.F., degrees of freedom
S.D., standard deviation
TK, Tukey-Kramer
N.A., not applicable
*, p<.05; **, p<.01

Table D3.4. Biomarker Differences Across Clusters in the MS Dataset for k=4

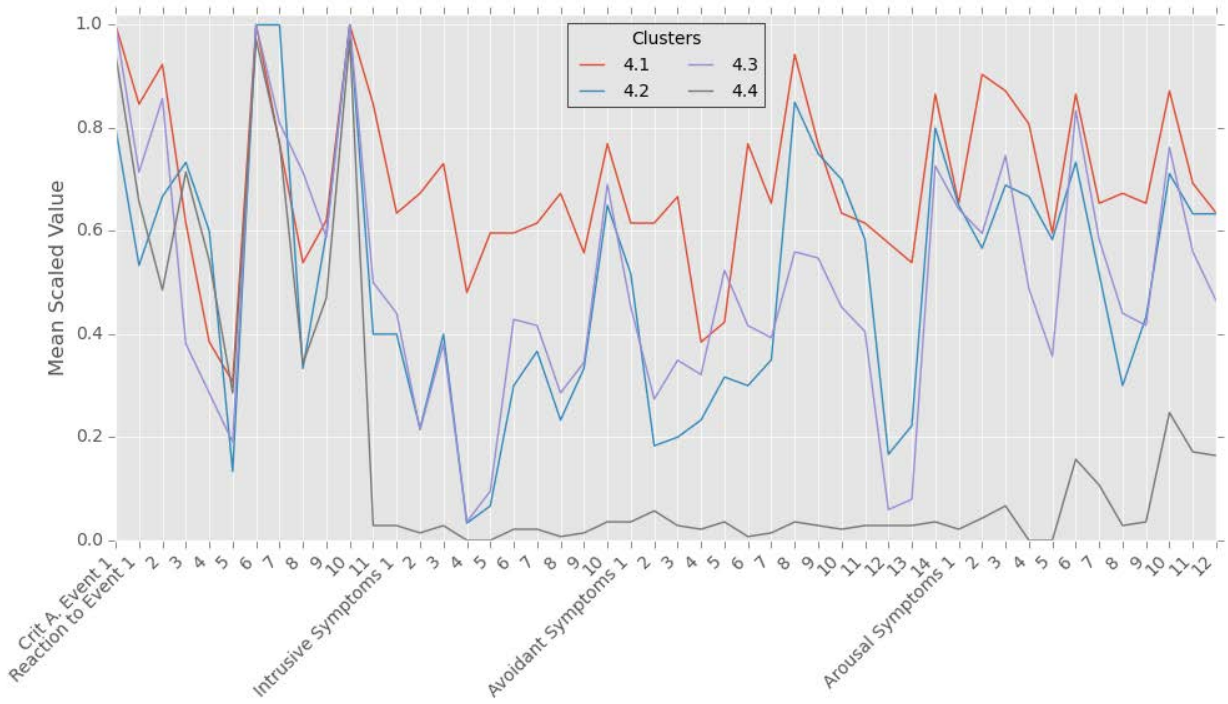

Figure D3.4a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=4)

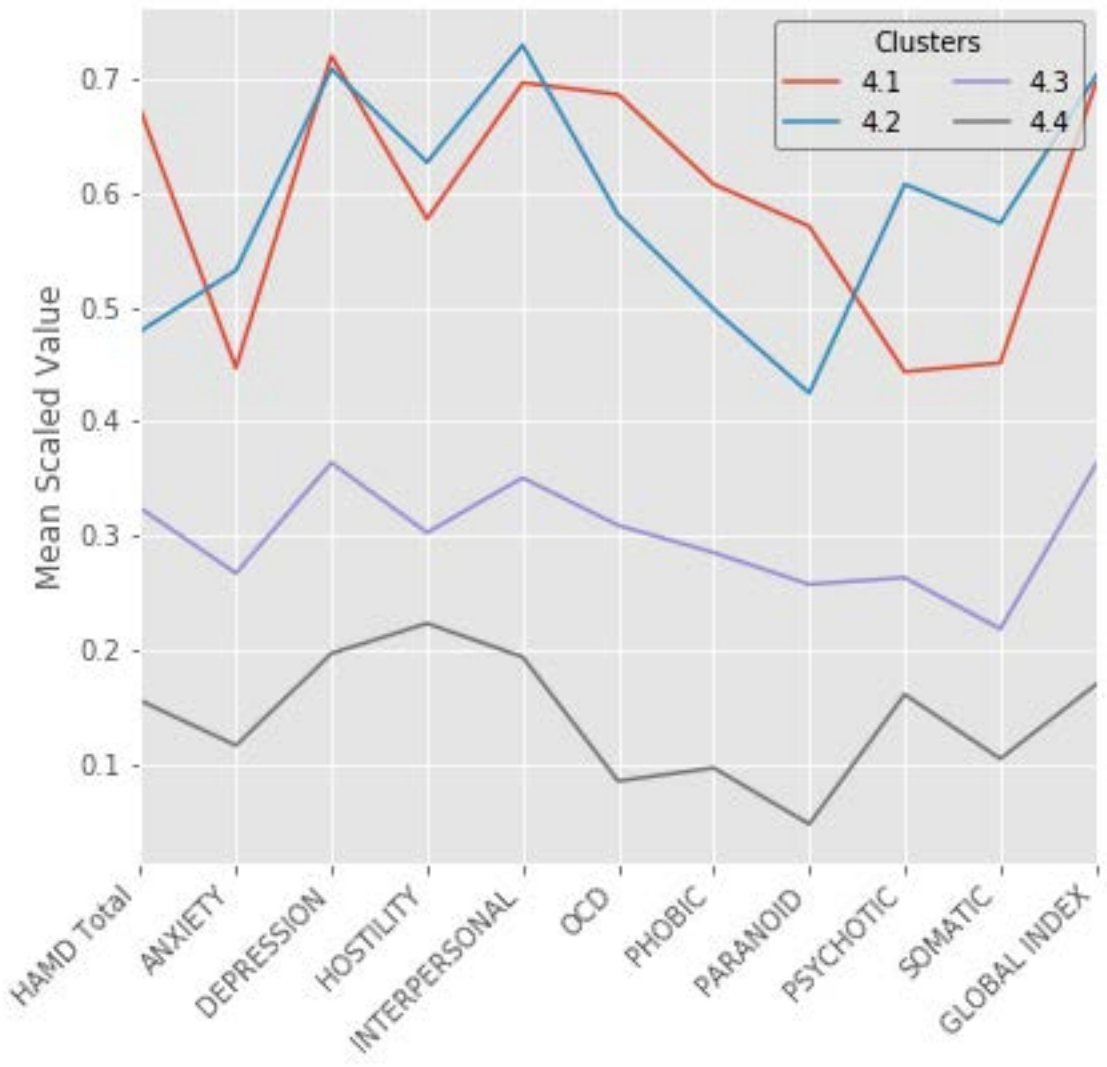

Figure D3.4b Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 (k=4)

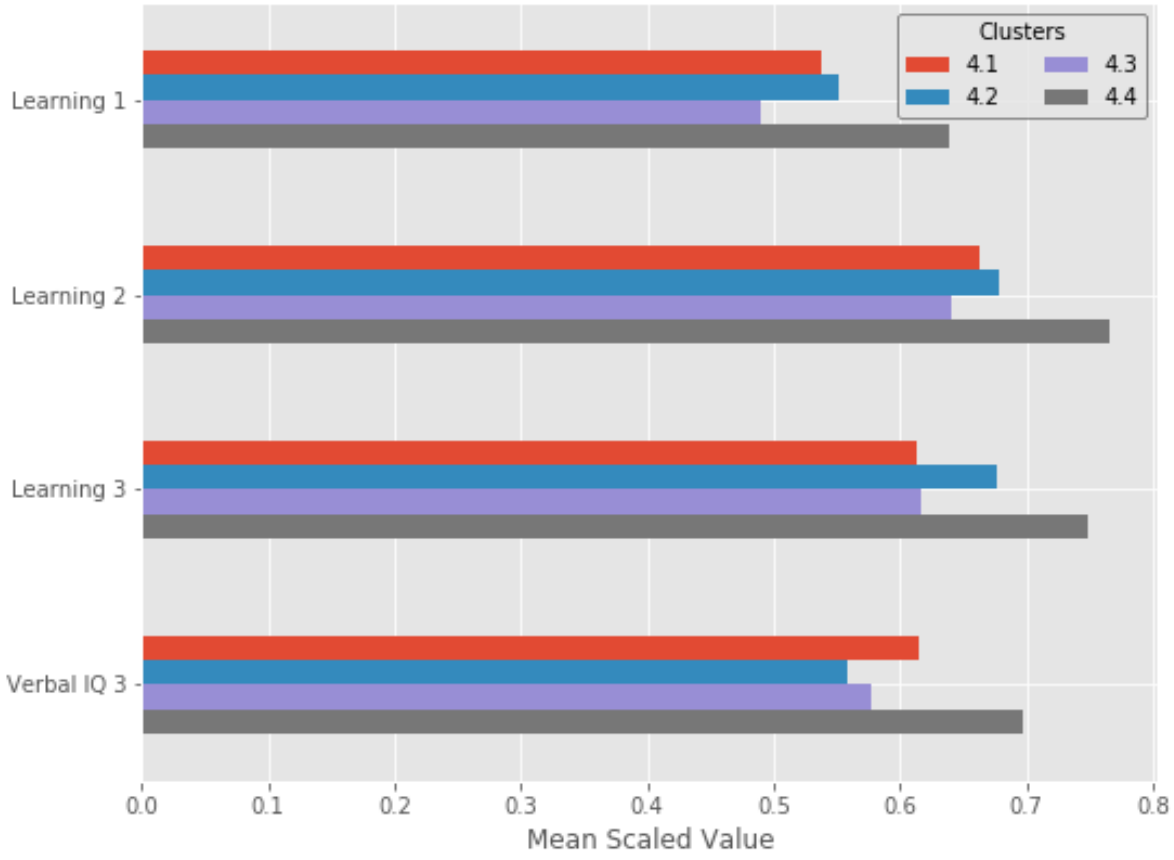

Figure D3.4c. Neuropsychiatric Markers with Significant Differences Across Clusters (k=4)

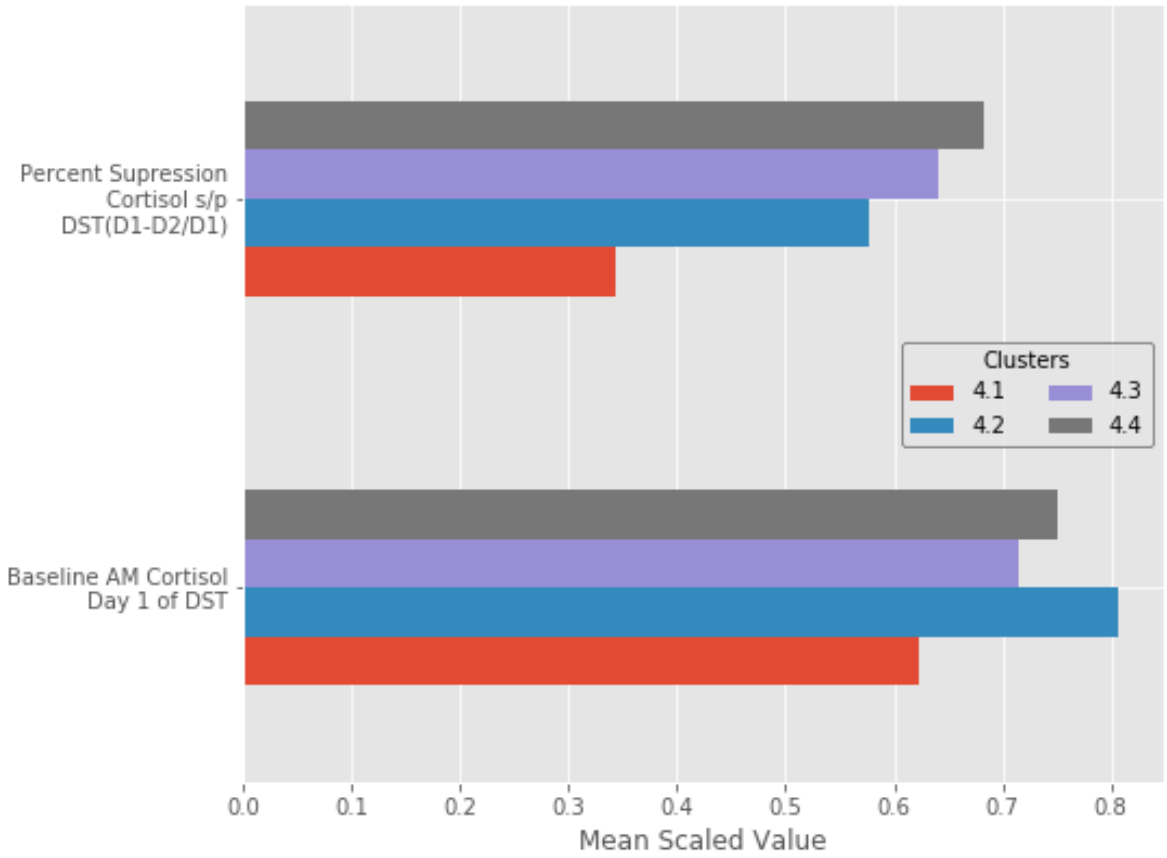

Figure D3.4d. Endocrine Markers with Significant Differences Across Clusters (k=4)

.

Test	Cluster 1 Mean	S.D.	Cluster 5.9 Mean	S.D.	Cluster 5.2 Mean	S.D.	Cluster 4 Mean	S.D.	Cluster 5 Mean	S.D.	ANCOV A P-value	Sig	F	Post Hoc Cluster Difference TK
Neuropsychiatric (D.F.=4,77)														
Learning 1	5.92E+00	1.91E+00	6.07E+00	1.44E+00	5.91E+00	1.72E+00	6.97E+00	1.28E+00	9.00E+00	N.A.	6.54E-09	**	9.86	4>3**
Learning 2	1.06E+01	9.22E+00	1.05E+01	2.67E+00	1.02E+01	2.17E+00	1.25E+01	2.24E+00	1.10E+01	N.A.	9.96E-02	*	2.76	4>3*
Learning 3	4.25E+01	1.22E+01	4.67E+01	9.06E+00	4.25E+01	9.95E+00	5.16E+01	8.95E+00	5.40E+01	N.A.	4.41E-03	**	4.13	4>1*,4>3
Verbal IQ 3	1.60E+01	9.76E+00	1.45E+01	9.54E+00	1.50E+01	9.32E+00	1.81E+01	9.98E+00	1.80E+01	N.A.	8.59E-03	**	9.67	4>1*,14>3*
Cortisol Measures (D.F.=4,63)														
Baseline AM Cortisol Day 1 of DST	2.89E+00	4.70E-01	9.75E+00	9.90E-01	9.92E+00	7.90E-01	9.97E+00	7.10E-01	9.92E+00	N.A.	1.29E-02	*	9.45	2>1**, 4>1*
Percent Suppression Baseline Cortisol in DST(D1-D2/D1)	4.41E+00	7.51E+00	1.62E+01	1.94E+01	1.54E+01	1.98E+01	1.50E+01	1.19E+01	1.55E+01	N.A.	9.80E-02	*	2.7	4>1*

D.F., degrees of freedom
S.D., standard deviation
TK, Tukey-Kramer
N.A., not applicable
*, p<.05; **, p<.01; ***, p<.001

Table D3.5. Biomarker Differences Across Clusters in the MS Dataset for k=5

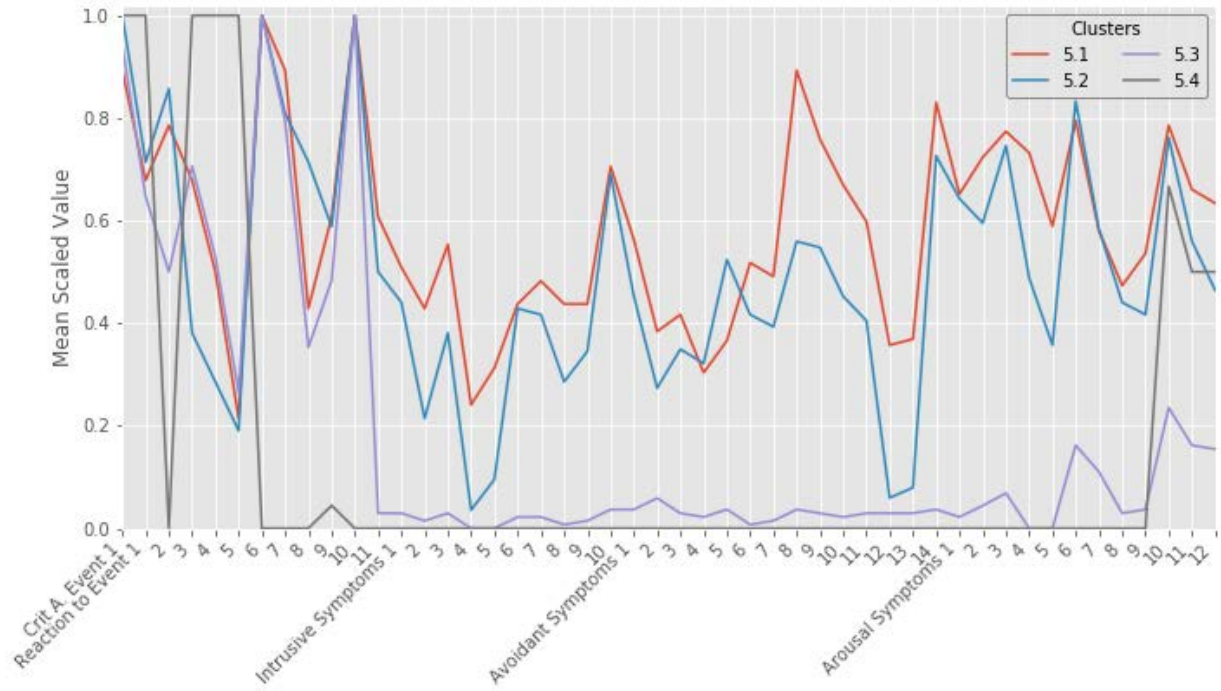

Figure D3.5a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=5)

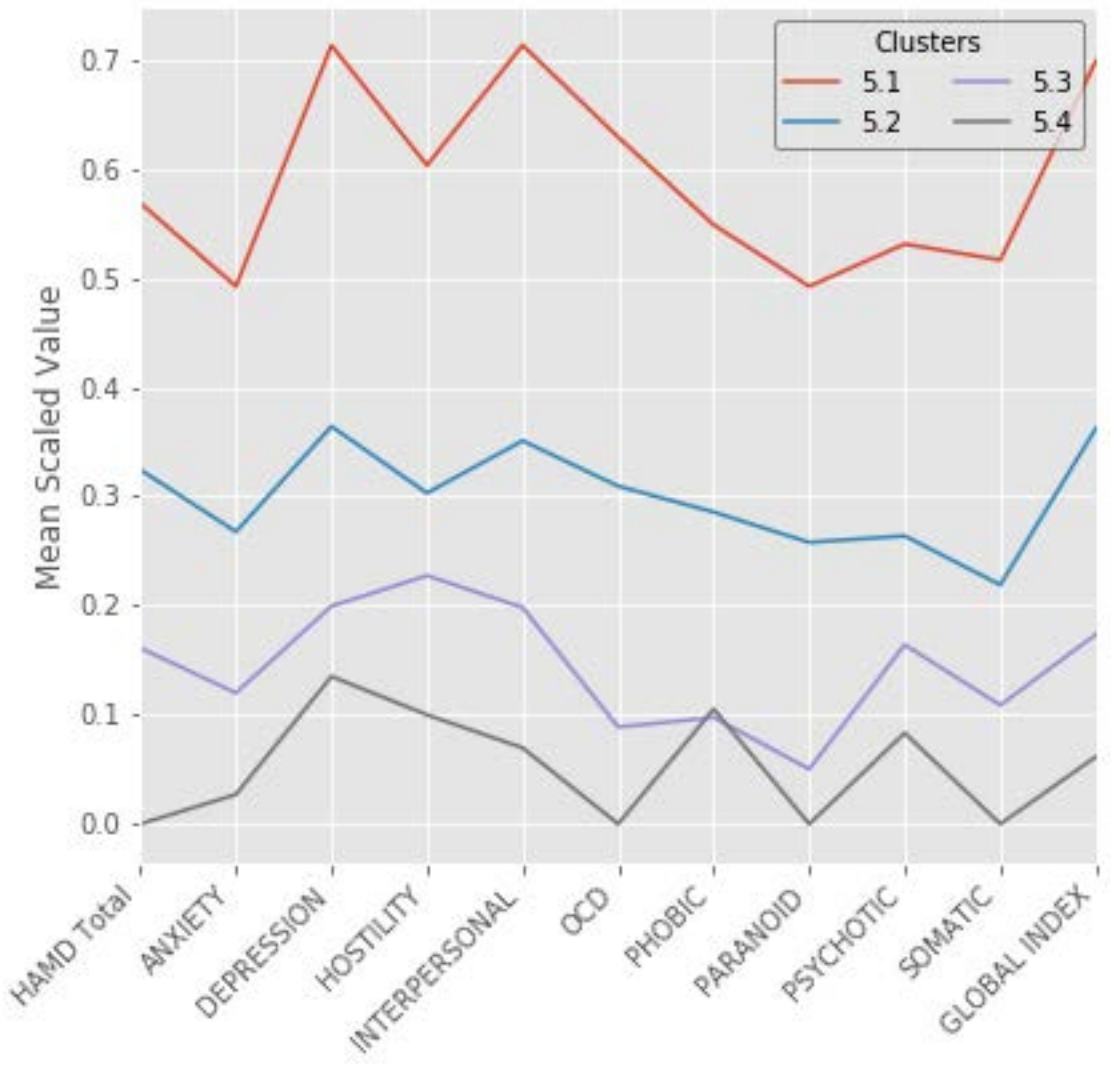

Figure D3.5b. Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 (k=5)

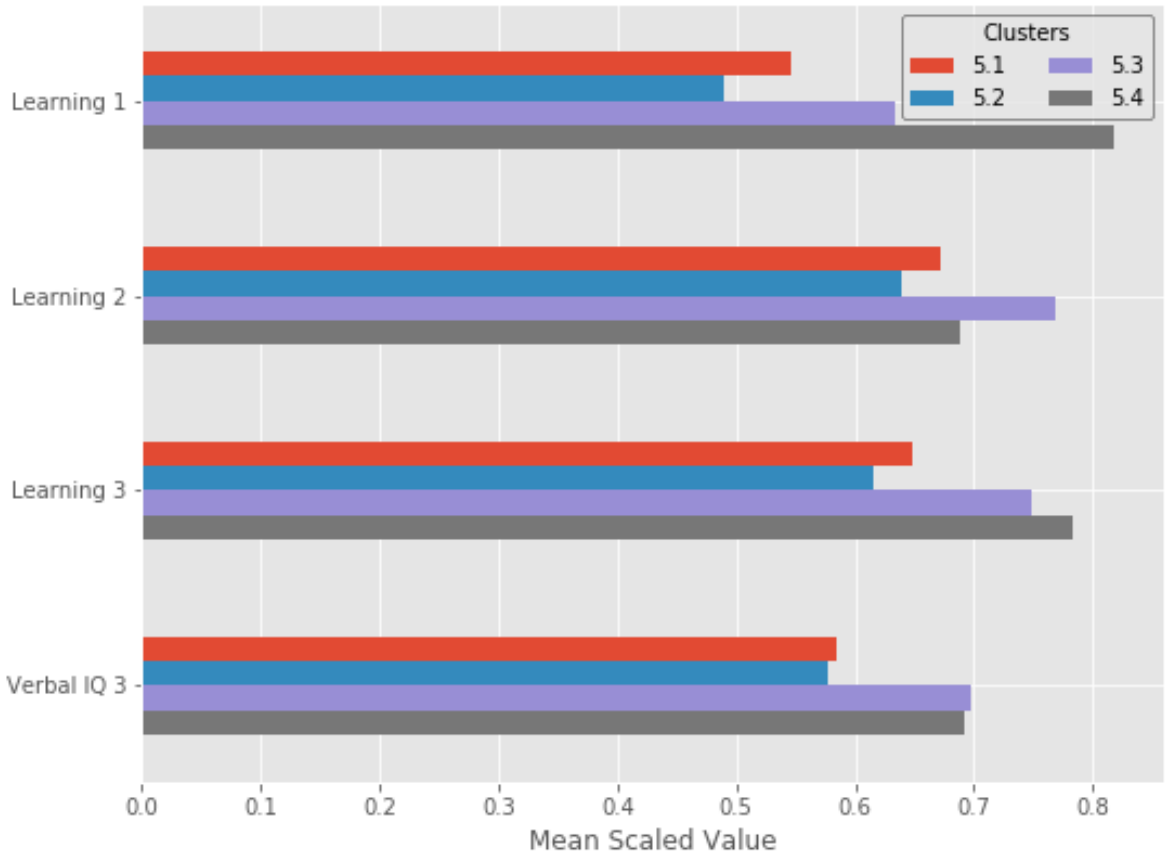

Figure D3.5c. Neuropsychiatric Markers with Significant Differences Across Clusters (k=5)

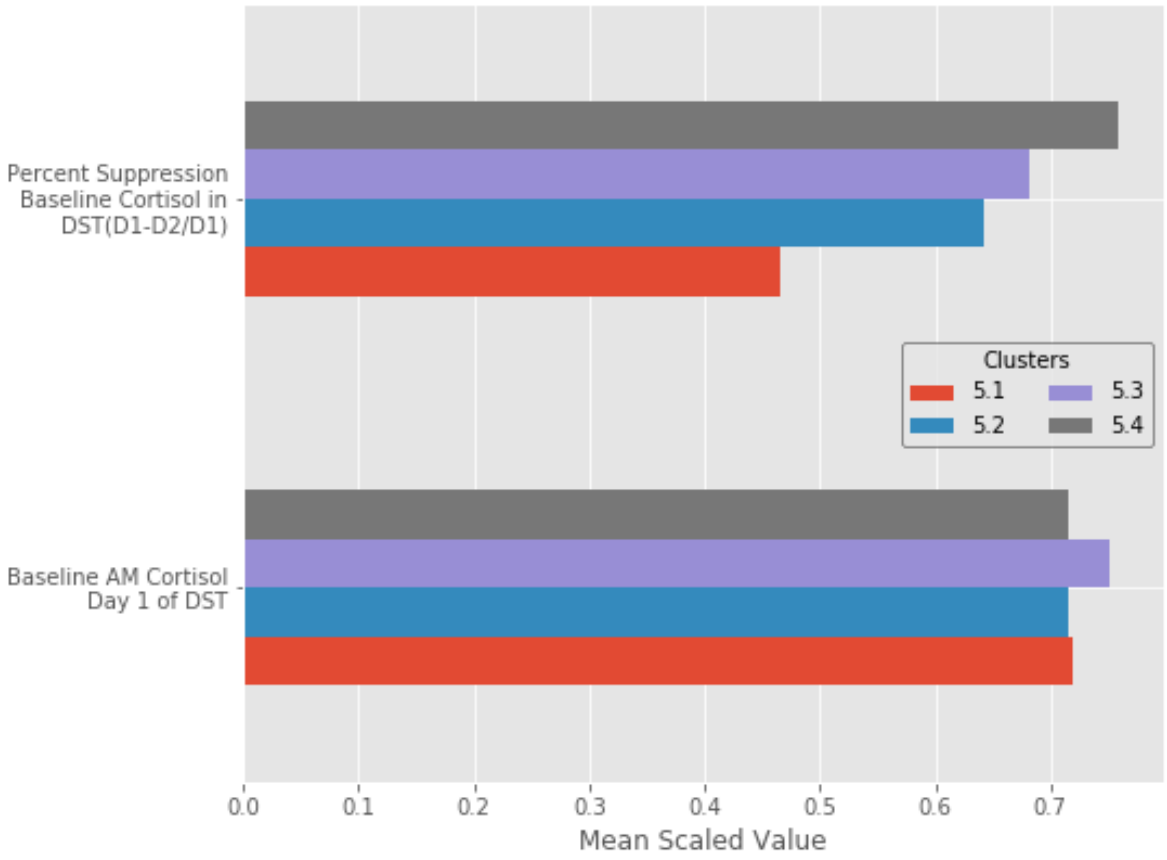

Figure D3.5d. Endocrine Markers with Significant Differences Across Clusters (k=5)

APPENDIX D4: Cluster Analyses of Combined Summary (CS) Dataset for k=2, k=3, k=4, and k=5 Solutions, with Symptom Profiles and Biomarker Differences

To examine the relationship between self-report symptom profiles and clinical and biological markers in the entire dataset, the Gulf War and MIRECC datasets were merged to form the Combined Summary (CS) dataset. The CS dataset includes all individuals who participated in the previous Gulf War and MIRECC analyses, clustered using the full set of 61 features from the CAPS, and the summary symptoms for the Ham-D and the SCL-90, as described in the analyses of the GS and MS datasets. There were 309 individuals in this dataset, 77 with PTSD, 41 with MDD, and 95 with alcohol dependence.

This appendix describes in detail the complete findings of the cluster analyses using the CS dataset for the k=2, k=3, and k=5 solutions. The main text in Chapter 6 describes the findings in the k=4 solution.

With the two-cluster solution, 75 out of 77 of the individuals with PTSD diagnoses were in Cluster 2.1, and these individuals accounted for 80 percent of the cluster (75/94). As expected, Cluster 2.1 had a significantly higher symptom profiles across all clinical inventories (Figures D4.2a and D4.2b). The individuals in Cluster 2.1 had a significantly greater proportion of alcohol dependence ($\chi 2=12.7$, $p<.001$) than individuals in Cluster 2.2, but there was no difference in the incidence of childhood trauma (Table D4.1).

As can be seen in Table D4.2, Cluster 2.2 has several biomarker domains with significantly different values from those in Cluster 2.1. The significance of these differences is generally greater than in the previous analyses, likely because this dataset is the largest. Cluster 2.1 performed worse across all neuropsychiatric domains (Table D4.2, Figure D4.2d), with significant differences in Learning, Executive Functioning, and Verbal IQ domains ($p<.05$,

p<.01, p<.001, p<0.0001). Additionally, there was a general pattern of smaller cortex volumes and greater white matter volumes in Cluster 2.1 than in Cluster 2.2 (p<.05, p<.01; Table D4.2, Figure D4.2d), with bilateral frontal and temporal cortices, and bilateral frontal, parietal, occipital, and temporal cortices all following this pattern. Finally, baseline AM cortisol was lower in Cluster 2.1 than in Cluster 2.2 (p<.01), as was percent suppression of cortisol following the DST (p<.001; Table D4.2, Figure D4.2e).

Table D4.1. Biomarker Differences Across Clusters in the CS Dataset for k=2

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	ANCOVA P-value	Sig	F	Post-Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=1,299)								
Learning 1	5.78E+00	1.73E+00	6.28E+00	1.89E+00	2.87E-02	*	4.83	2>1*
Learning 2	1.10E+01	2.79E+00	1.24E+01	2.54E+00	3.86E-05	****	17.46	2>1****
Learning 3	4.56E+01	1.07E+01	5.10E+01	9.76E+00	1.26E-05	****	19.43	2>1****
Executive Functioning 1	9.23E+00	3.07E+00	1.08E+01	3.03E+00	3.81E-05	****	17.49	2>1****
Executive Functioning 2	1.05E+01	2.77E+00	1.19E+01	2.67E+00	8.41E-05	****	15.89	2>1***
Memory 1	9.65E+00	3.42E+00	1.13E+01	3.02E+00	3.00E-05	****	17.97	2>1****
Memory 2	1.05E+01	3.06E+00	1.20E+01	2.80E+00	4.48E-05	****	17.16	2>1****
Performance IQ 1	6.36E+01	1.28E+01	6.79E+01	1.51E+01	1.08E-02	*	6.57	2>1*
Performance IQ 2	3.78E+01	1.15E+01	4.31E+01	1.95E+00	1.98E-04	***	14.2	2>1***
Verbal IQ 3	1.55E+01	3.80E+00	1.72E+01	4.26E+00	1.36E-03	**	10.46	2>1**
Imaging Volumes (D.F.=1,165)								
Right Frontal Cortex	8.22E-02	8.70E-03	8.46E-02	5.50E-03	2.07E-03	**	9.713	2>1*
Left Frontal Cortex	8.20E-02	8.30E-03	8.43E-02	5.70E-03	3.08E-03	**	8.951	2>1*
Right Temporal Cortex	4.97E-02	5.70E-03	5.18E-02	3.80E-03	2.29E-04	***	14.03	2>1**
Left Temporal Cortex	4.96E-02	5.50E-03	5.15E-02	3.80E-03	7.30E-04	***	11.74	2>1**
Right Frontal White Matter	9.42E-02	1.32E-02	8.93E-02	7.30E-03	4.26E-04	***	12.79	1>2***
Left Frontal White Matter	9.34E-02	1.39E-02	8.89E-02	7.60E-03	1.56E-03	**	10.26	1>2**
Right Parietal White Matter	4.73E-02	7.60E-03	4.51E-02	3.70E-03	5.77E-03	**	7.77	1>2**
Left Parietal White Matter	4.57E-02	6.90E-03	4.35E-02	3.40E-03	1.42E-03	**	10.44	1>2**
Right Occipital White Matter	1.76E-02	2.40E-03	1.67E-02	2.00E-03	2.94E-03	**	9.05	1>2**
Left Occipital White Matter	1.78E-02	2.70E-03	1.68E-02	2.20E-03	3.04E-03	**	8.98	1>2**
Right Temporal White Matter	3.71E-02	7.90E-03	3.48E-02	3.50E-03	1.81E-03	**	9.97	1>2**
Left Temporal White Matter	3.65E-02	8.20E-03	3.46E-02	3.30E-03	1.27E-02	*	6.31	1>2*
Right Parietal CSF	2.02E-02	4.80E-03	2.24E-02	5.40E-03	2.25E-03	**	9.56	2>1**
Left Parietal CSF	1.95E-02	4.90E-03	2.13E-02	4.90E-03	4.44E-03	**	8.26	2>1**
Left Occipital CSF	4.35E-03	1.30E-03	4.89E-03	1.40E-03	5.33E-03	**	7.92	2>1**
Right Temporal CSF	1.42E-02	2.96E-03	1.50E-02	2.86E-03	2.87E-02	*	4.85	N.S.
Cortisol Measures (D.F.=1,209)								
Baseline AM Cortisol Day 1 of DST	3.21E+00	6.80E-01	3.38E+00	4.90E-01	3.64E-02	*	4.37	2>1*
Percent Suppression Baseline Cortisol in DST(D1-D2/D1)	6.97E+01	3.16E+01	8.32E+01	2.09E+01	2.95E-04	***	13.56	2>1***

D.F.: degrees of freedom
S.D.: standard deviation
N.S.: not significant
TK: Tukey-Kramer
*, p<.05; **, p<.01; ***, p<.001; ****, p<.0001

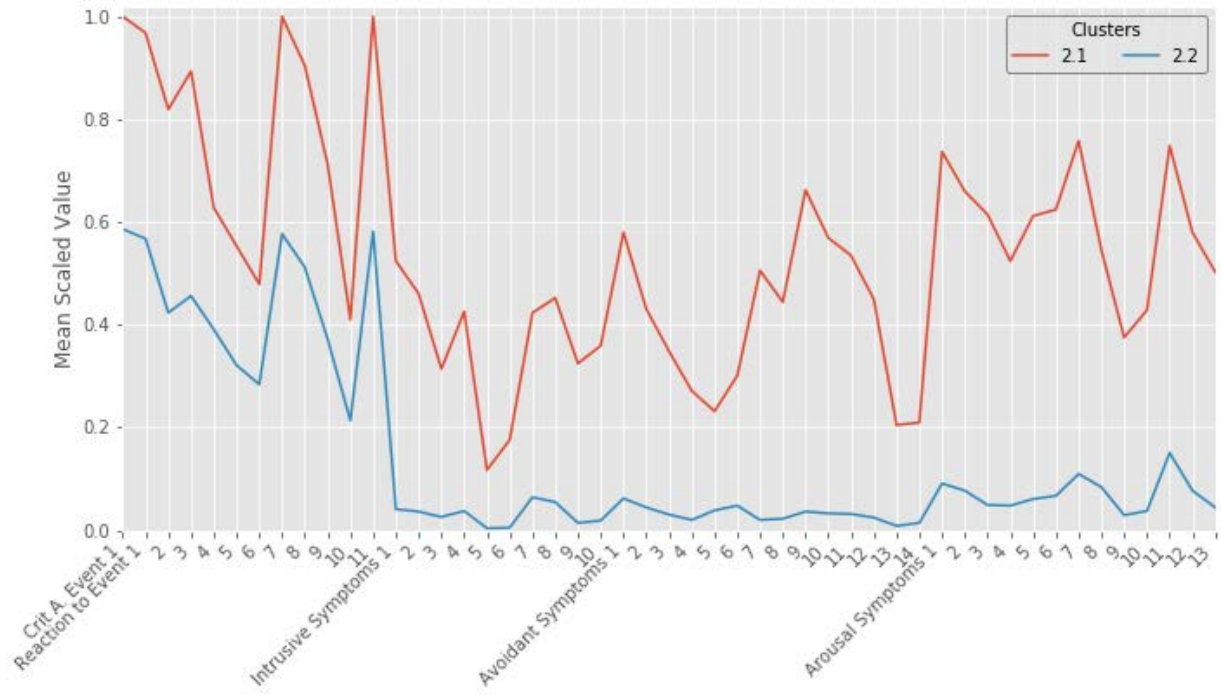

Figure D4.1a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=2)

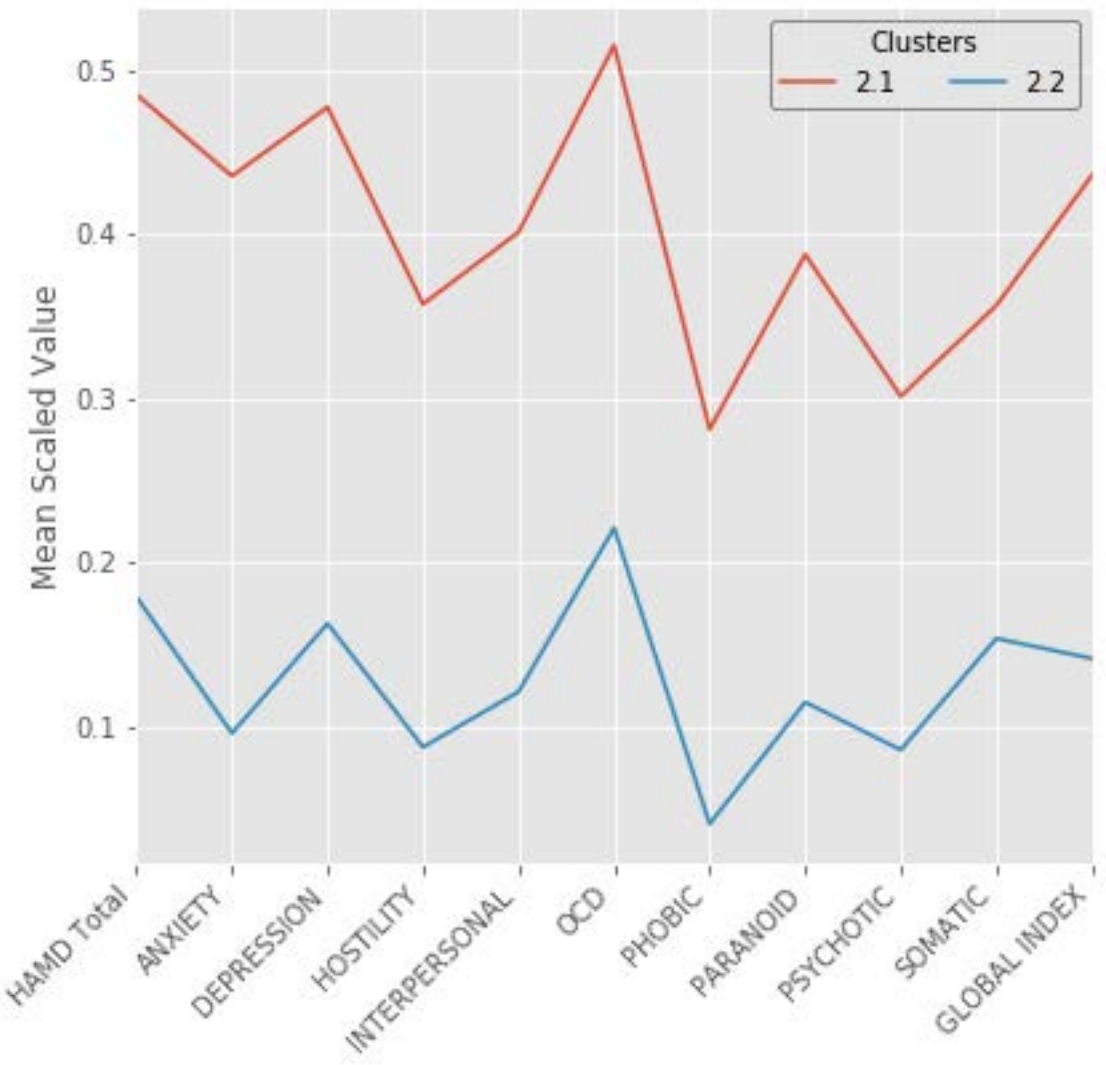

Figure D4.1b. Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 (k=2)

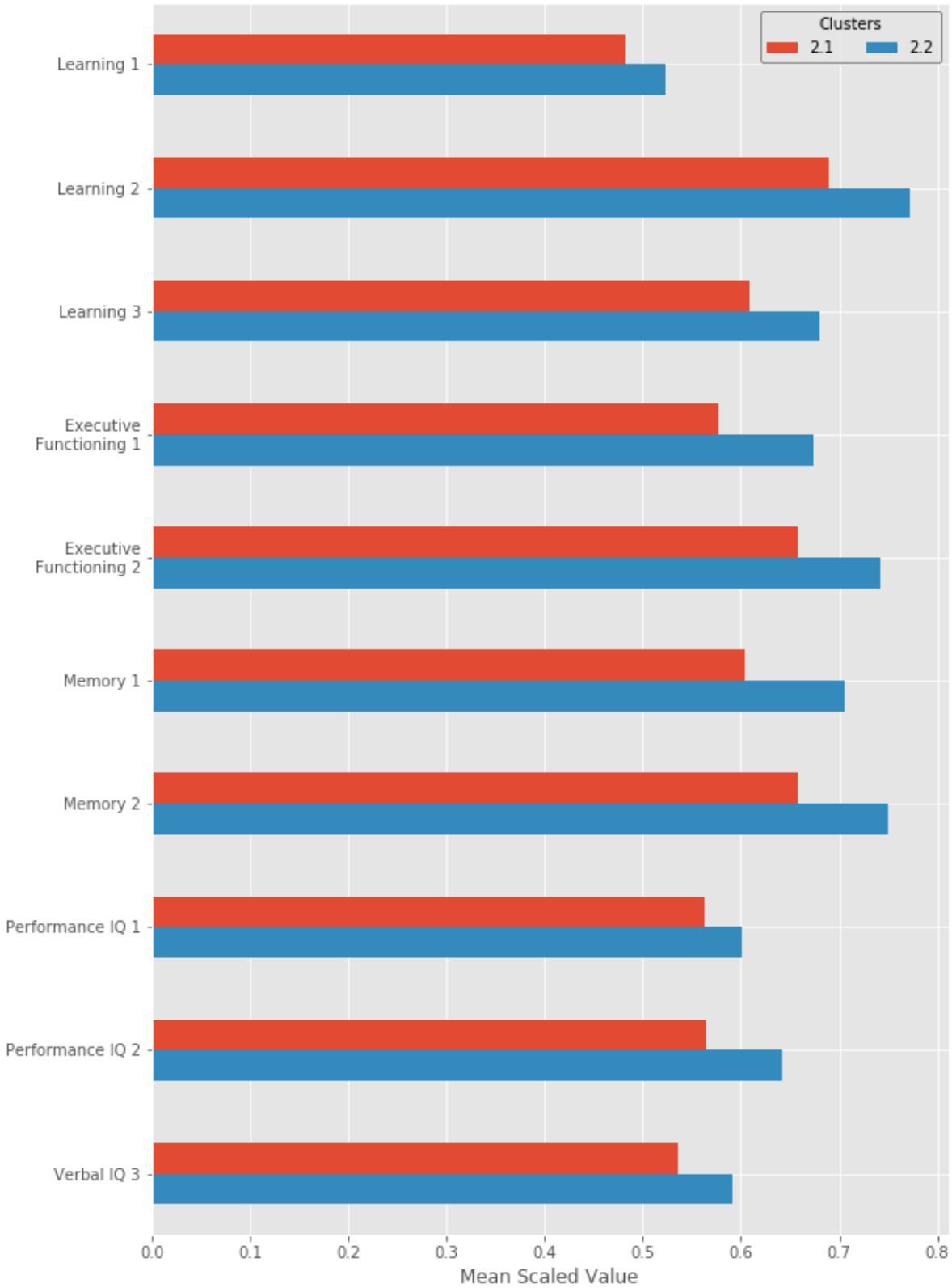

Figure D4.1c. Neuropsychiatric Markers with Significant Differences Across Clusters (k=2)

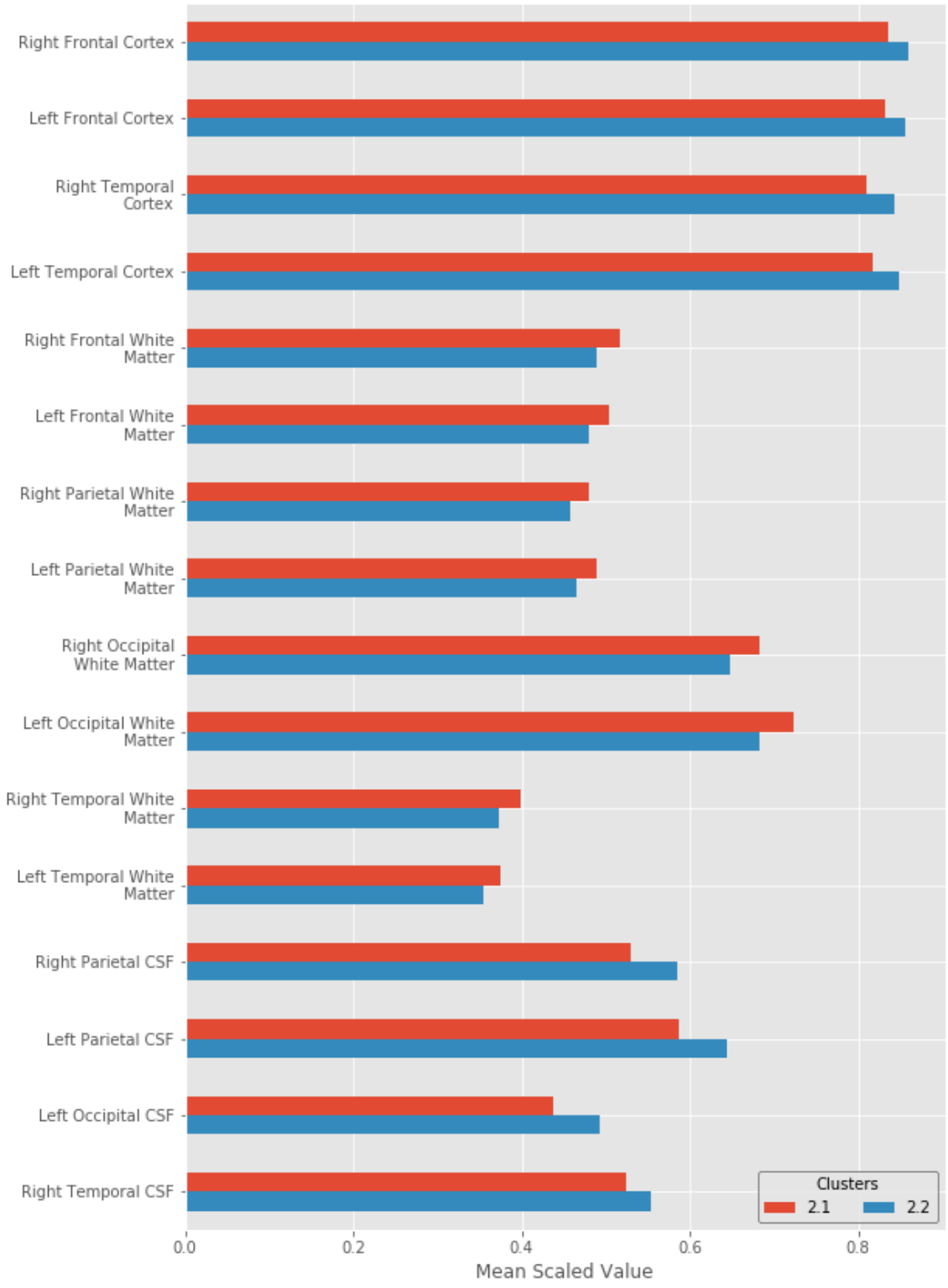

Figure D4.1d. Imaging Markers with Significant Differences Across Clusters (k=2)

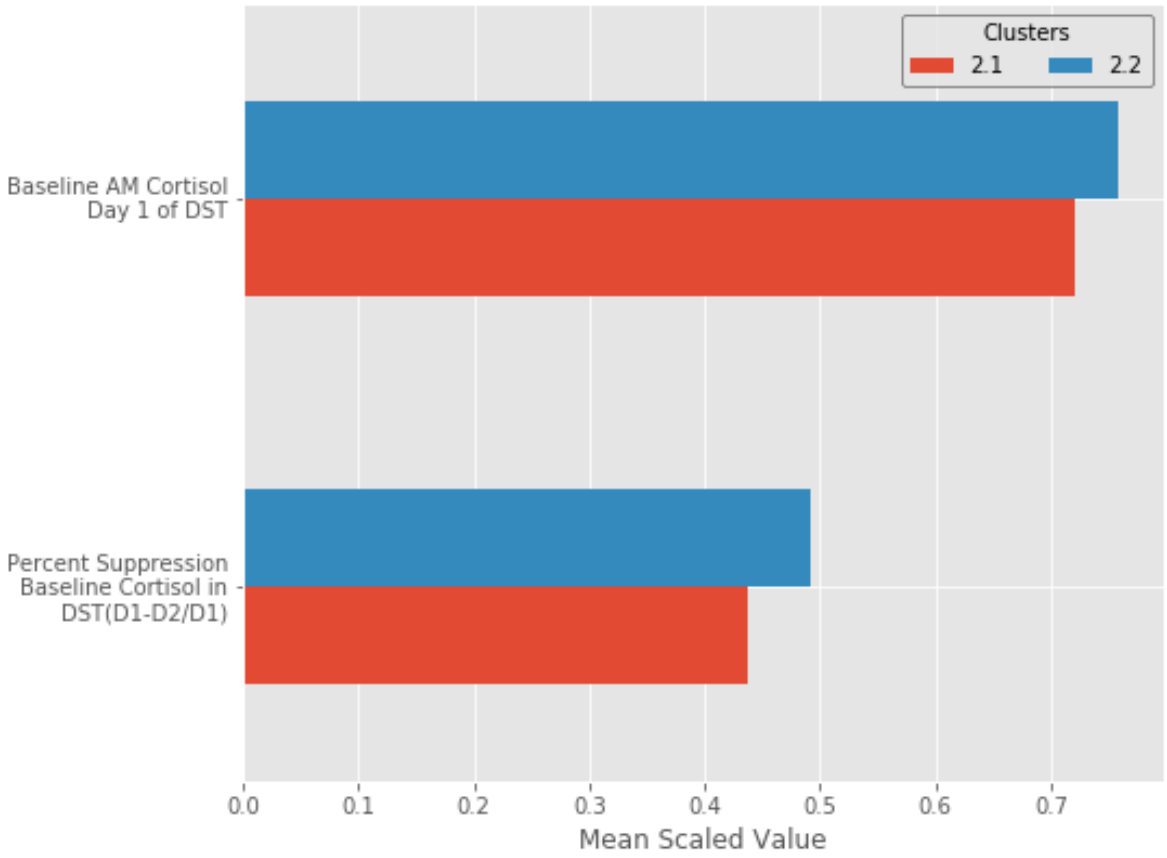

Figure D4.1e. Endocrine Markers with Significant Differences Across Clusters (k=2)

The three-cluster solution also performed well, with an ARI of 0.8 and 100 percent cluster stability during bootstrap analysis. Table D4.1 shows that Cluster 3.1 differed only slightly from Cluster 2.1; Cluster 3.1 also comprised 80 percent of the individuals with PTSD, and >85 percent of this cluster had PTSD diagnoses. On average, individuals in Cluster 3.1 had three DSM diagnoses, whereas those in Cluster 3.2 had 1.3 and Cluster 3.3 had 0.7 other DSM diagnoses, respectively (Table D4.3). The symptom profile for Cluster 3.1 was consistent to that of Cluster 2.1, with higher levels of distress across all clinical areas compared to Clusters 3.2 and 3.3 (Figures D4.3a–b). The incidence of alcohol dependence across all three clusters differed significantly ($\chi 2=12.5$, $p<.01$), with the percentage of individuals with alcohol dependence being the greatest in Cluster 3.1 and least in Cluster 3.3, decreasing as the symptom profiles lowered (Table D4.1). This decreasing pattern as the cluster symptom profile decreased was also seen in the incidence of childhood trauma ($\chi 2=14.8$, $p<.001$; Table D4.1).

Individuals in Cluster 3.1, on average, performed significantly worse across all neuropsychiatric domains than individuals in Clusters 3.2, 3.3, or the combination of both clusters ($p<.05$, $p<.01$, $p<.001$, $p<.0001$; Table D4.3, Figure D4.3c). As with Cluster 2.1 in the two-cluster solution, individuals in Cluster 3.1 had significantly smaller bilateral frontal and temporal cortices than those in either Cluster 3.2 or Cluster 3.3 ($p<.05$, $p<.01$; Table D4.3, Figure D4.3d). However, Cluster 3.1 also had significantly larger white matter volumes bilaterally across all lobes than Cluster 3.3 ($p<.05$, $p<.01$, $p<.001$; Table D4.3, Figure D4.3d). Finally, baseline cortisol was essentially the same for Clusters 3.1 and 3.2, but was higher in Cluster 3.3 ($p<.01$; Figure D4.3e).

Test	Cluster 1 Mean	S.D.	Cluster 2 Mean	S.D.	Cluster 3 Mean	S.D.	ANCOVA P-value	Sig	F	Post Hoc Cluster Difference[TK]
Neuropsychiatric (D.F.=2,29)										
Executive Functioning 1	9.07E+00	9.04	1.08E+01	2.96	1.08E+01	9.12	9.88E-05	****	10.57	3>1***,2>1***
Executive Functioning 2	1.04E+01	2.75	1.20E+01	2.43	1.17E+01	2.97	4.75E-05	****	10.29	3>1**,2>1****
Learning 1	5.74E+00	1.72	6.97E+00	1.91	6.15E+00	1.83	4.71E-02	*	3.09	2>1*
Learning 2	1.05E+01	2.72	1.24E+01	2.96	1.23E+01	2.82	2.90E-05	****	10.82	3>1**,2>1****
Learning 3	4.45E+01	10.99	5.16E+01	9.55	5.06E+01	10.23	9.87E-06	****	12.99	3>1***,2>1*****
Memory 1	9.44E+00	9.99	1.14E+01	2.9	1.15E+01	9.19	1.38E-05	****	11.63	3>1***,2>1****
Memory 2	1.09E+01	9.05	1.21E+01	2.73	1.20E+01	2.86	2.00E-05	****	11.22	3>1***,2>1****
Performance IQ 1	6.90E+01	12.94	6.67E+01	15.92	7.00E+01	13.1	2.84E-03	**	5.98	3>1**
Performance IQ 2	9.71E+01	11.29	4.22E+01	12.01	4.48E+01	11.68	2.12E-05	****	11.16	3>1****,2>1**
Verbal IQ 3	1.59E+01	9.65	1.70E+01	4.96	1.75E+01	4.14	7.74E-04	***	7.94	3>1**,2>1**
Imaging Volumes (D.F.=2,22)										
Right Frontal Cortex	8.20E-02	0.0089	8.50E-02	0.0053	8.41E-02	0.0057	9.14E-03	**	5.92	2>1*
Left Frontal Cortex	8.18E-02	0.0085	8.49E-02	0.0054	8.34E-02	0.0059	2.32E-03	**	6.34	2>1*
Right Temporal Cortex	4.94E-02	0.0057	5.18E-02	0.004	5.18E-02	0.0054	1.56E-04	***	9.12	3>1**,2>1**
Left Temporal Cortex	4.95E-02	0.0056	5.16E-02	0.004	5.15E-02	0.0055	1.01E-03	**	7.18	3>1*,2>1**
Right Frontal White Matter	9.43E-02	0.0196	9.04E-02	0.0072	8.78E-02	0.0071	5.22E-04	***	7.82	3>1***, 3>2*
Left Frontal White Matter	9.35E-02	0.0143	9.00E-02	0.0077	8.74E-02	0.0071	2.81E-03	**	6.03	3>1**
Right Parietal White Matter	4.72E-02	0.0078	4.55E-02	0.0036	4.48E-02	0.0039	2.77E-02	*	3.64	3>1*
Left Parietal White Matter	4.57E-02	0.007	4.35E-02	0.0034	4.35E-02	0.0033	9.56E-03	**	4.75	3>1*
Right Occipital White Matter	1.76E-02	0.0024	1.69E-02	0.0021	1.64E-02	0.0019	5.98E-03	**	5.24	3>1**
Left Occipital White Matter	1.78E-02	0.0027	1.72E-02	0.0022	1.63E-02	0.002	1.15E-03	**	6.94	3>1***
Right Temporal White Matter	9.72E-02	0.0081	9.54E-02	0.004	9.40E-02	0.0034	2.61E-03	**	6.11	3>1**
Left Temporal White Matter	9.65E-02	0.0085	9.50E-02	0.0036	9.40E-02	0.0027	2.75E-02	*	3.66	3>1*
Right Parietal CSF	2.08E-02	0.0049	2.11E-02	0.0054	2.40E-02	0.0048	3.20E-05	****	10.85	3>1***,3>2**
Left Parietal CSF	1.95E-02	0.005	2.02E-02	0.0049	2.29E-02	0.0046	4.31E-05	****	10.53	3>1***,3>2**
Right Occipital CSF	4.58E-03	0.0015	4.58E-03	0.0012	5.55E-03	0.0014	2.70E-05	****	11.03	3>1***,3>2*****
Left Occipital CSF	4.34E-03	0.0013	4.43E-03	0.0012	5.55E-03	0.0015	8.20E-08	****	17.58	3>1****,3>2****
Right Temporal CSF	1.43E-02	0.003	1.45E-02	0.0027	1.57E-02	0.0029	1.80E-03	**	6.51	3>1*,3>2*
Left Temporal CSF	1.32E-02	2.91E-03	1.38E-02	2.95E-03	1.44E-02	3.02E-03	1.40E-02	*	4.95	N.S.
Cortical Measures (D.F.=2,288)										
Percent Suppression Baseline Contrast in DS 1(D1-D2/D1)	7.05E+01	9.09E+01	8.00E+01	2.99E+01	8.63E+01	1.92E+01	2.26E-03	**	6.27	3>1**

D.F., degrees of freedom
S.D., standard deviation
TK, Tukey-Kramer N.
*, p<.05; **, p<.01; ***, p<.001; ****, p<.0001
N.S., not Significant

Table D4.2. Biomarker Differences Across Clusters in the CS Dataset for k=3

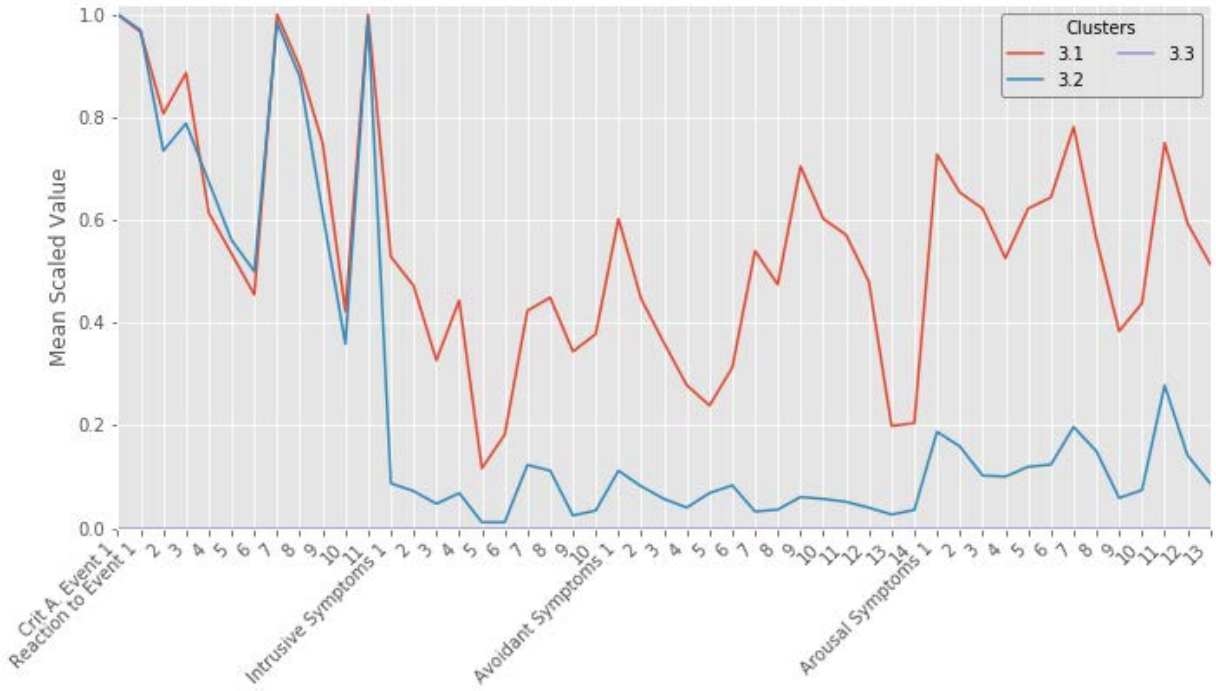

Figure D4.2a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=3)

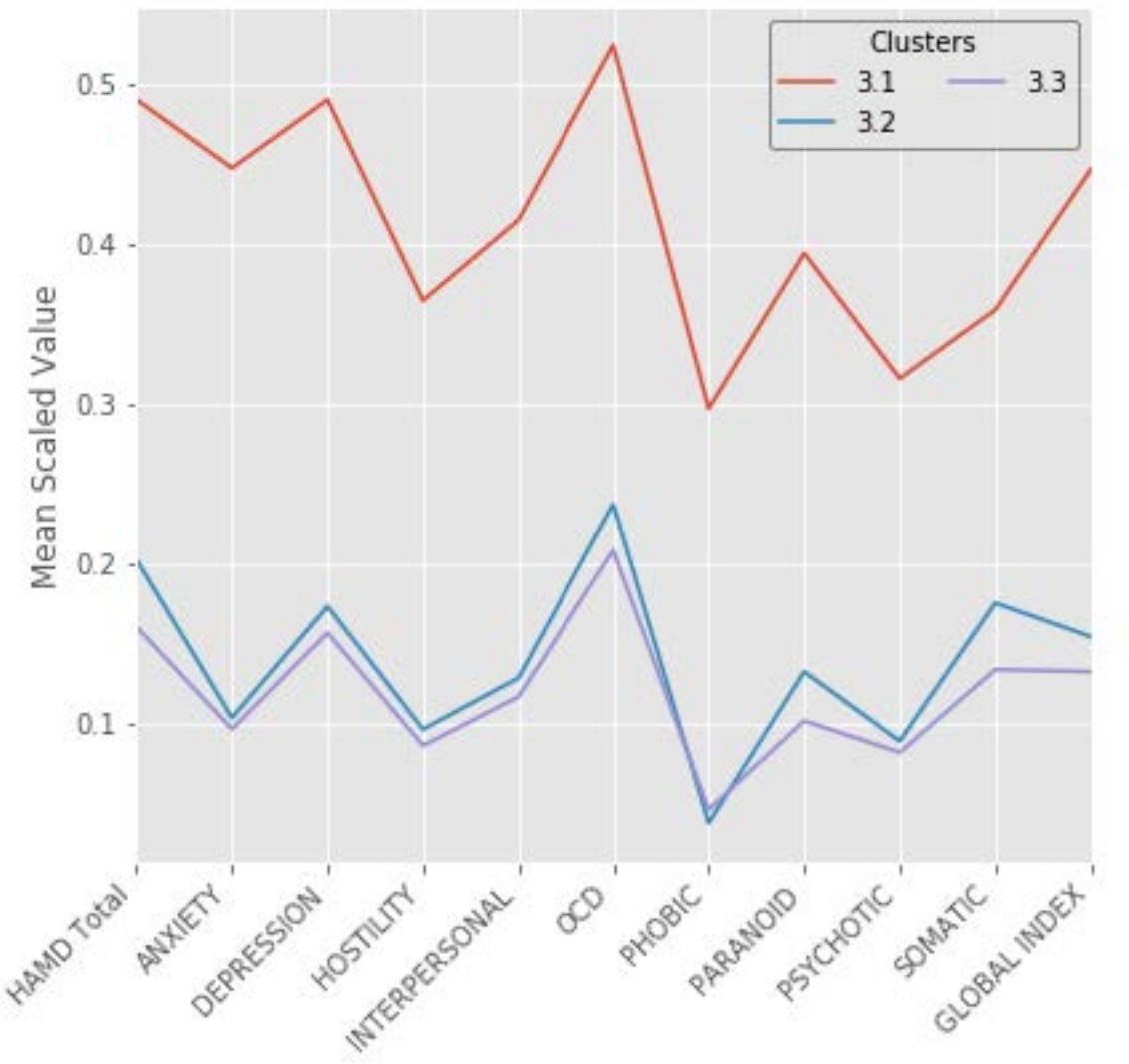

Figure D4.2b. Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 (k=3)

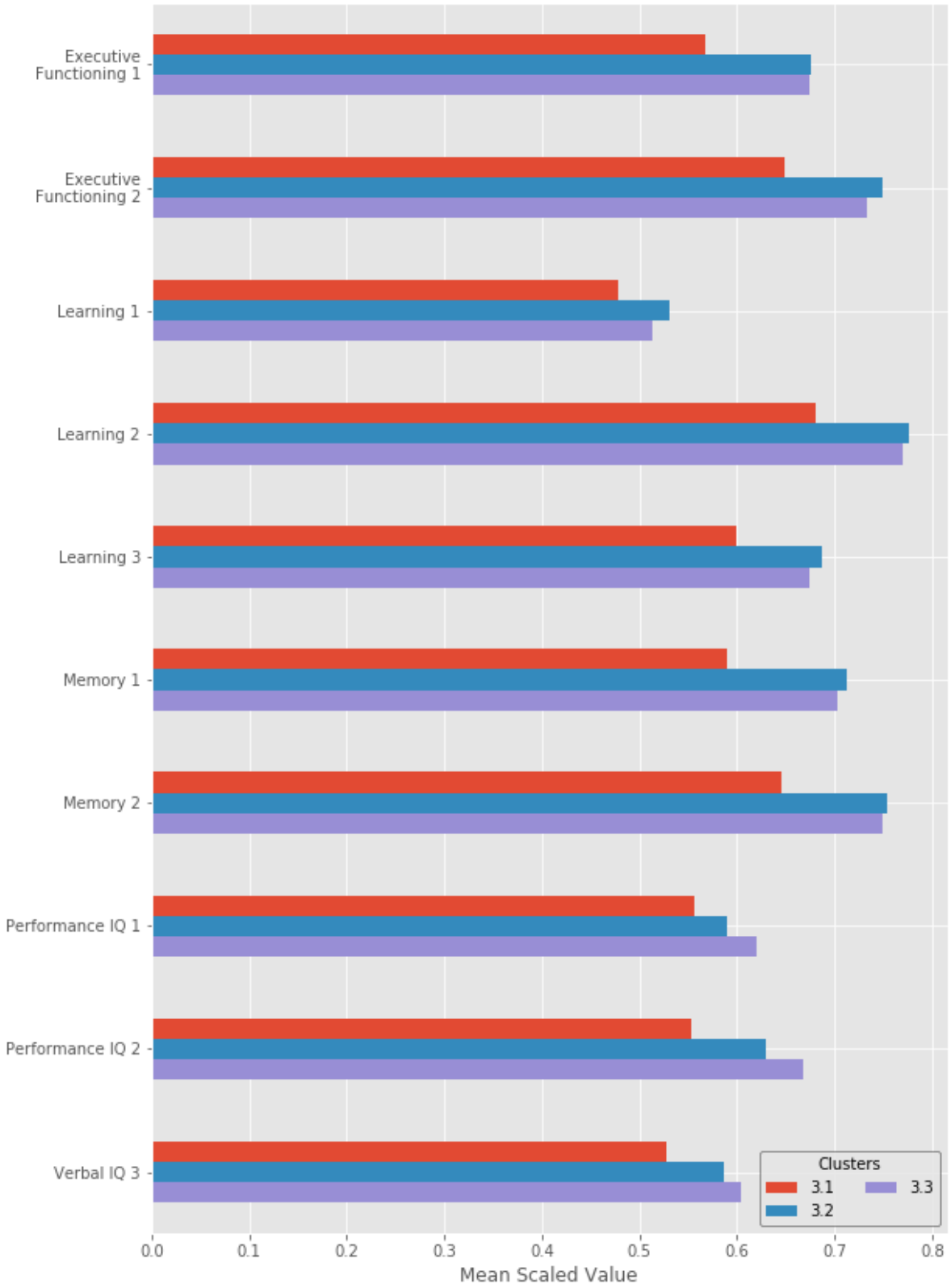

Figure D4.2c. Neuropsychiatric Markers with Significant Differences Across Clusters (k=3)

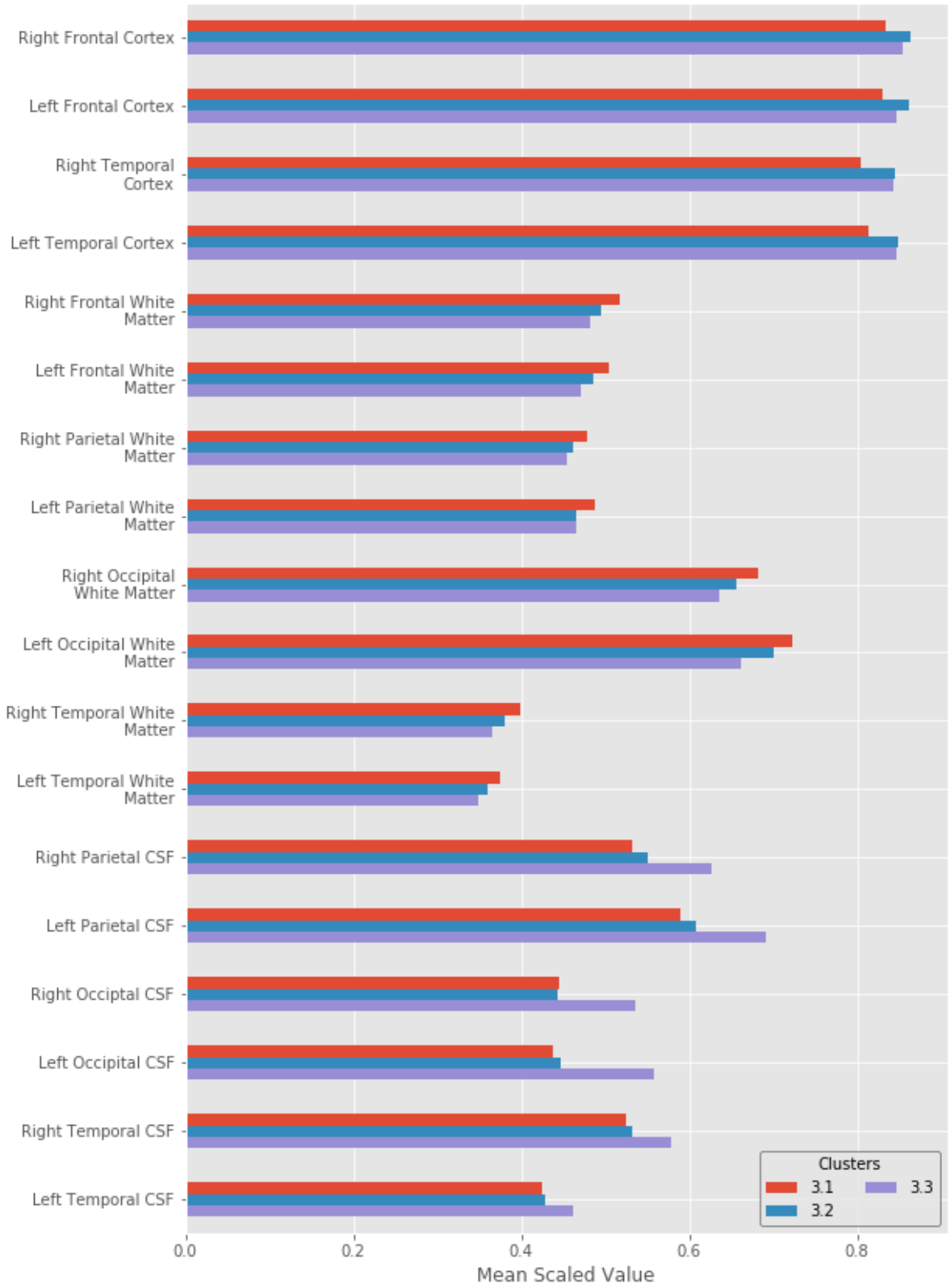

Figure D4.2d. Imaging Markers with Significant Differences Across Clusters (k=3)

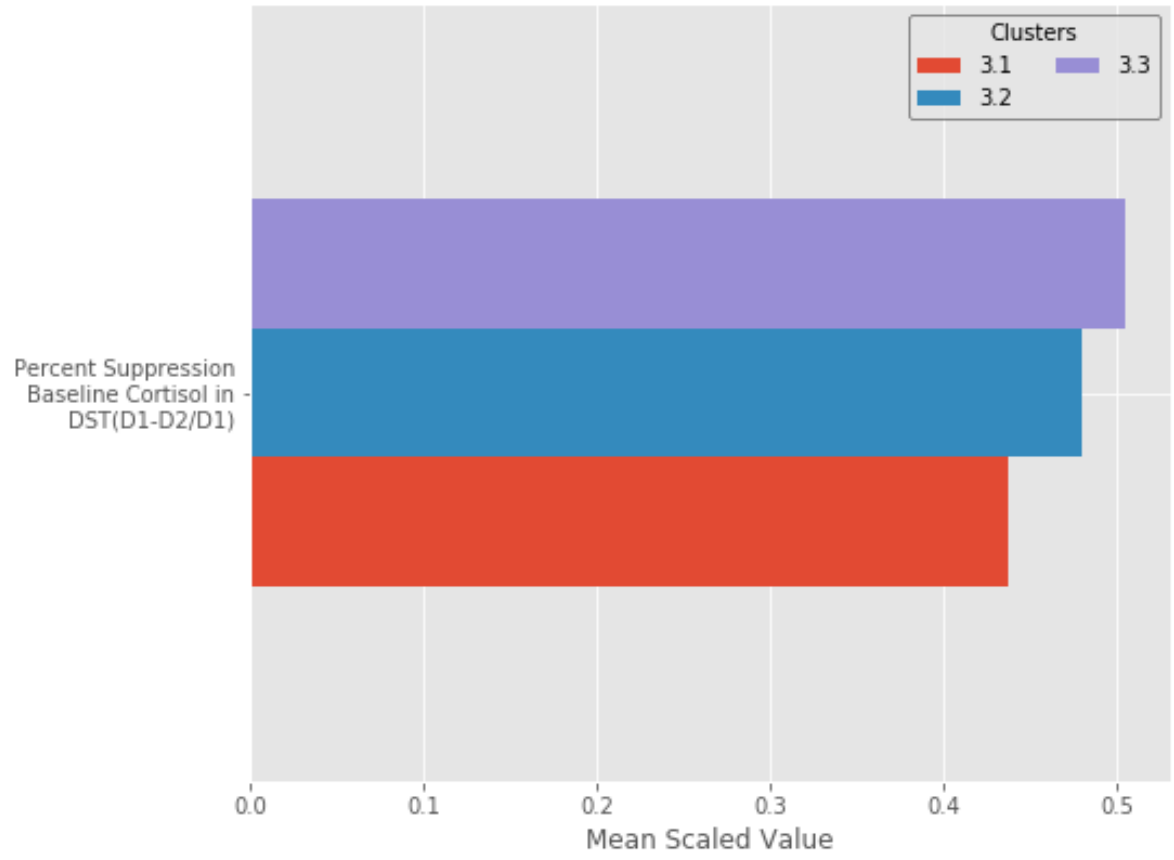

Figure D4.2e. Endocrine Markers with Significant Differences Across Clusters (k=3)

The five-cluster solution had a lower ARI (0.76) than the other solutions, and the bootstrap criteria did not meet the 80 percent cutoff for internal validity in two clusters, supporting the conclusion that this solution was not as robust as the previous three. However, there are some interesting findings in this solution as well. Three out of five clusters contained all of the individuals with PTSD. Cluster 5.1 was composed entirely (100 percent) of PTSD patients, with an average of 3.2 DSM diagnoses per individual (Table D4.1). Cluster 5.2 had 42 individuals with PTSD, composing 88 percent of the group, and an average of 2.9 DSM diagnoses per individual. Finally, Cluster 5.3 had the remaining 16 individuals with PTSD, making up 35 percent of the 46-person cluster. Cluster 5.3 had an average of 2.2 DSM diagnoses per individual (Table D4.1). Clusters 5.4 and 5.5 had no individuals with PTSD, and an average of 1.3 and 0.7 DSM diagnoses per individual, respectively (Table D4.1). Consistently, Clusters 5.1, 5.2, and 5.3 had higher levels of symptom distress (in descending order) than Clusters 5.5 and 5.4 (Figures D4.3a–b). The incidence of alcohol dependence across all five clusters differed significantly ($\chi 2=14.2$, $p<.0001$), with the percentage of individuals with alcohol dependence being the greatest in Cluster 5.2, followed by Clusters 5.3, 5.1, 5.4, and 5.5 (Table D4.1). The incidence of childhood trauma also significantly differed across the five clusters, with the highest incidence in Cluster 5.1, followed by 5.4, 5.2, 5.3, and 5.4 ($\chi 2=17.3$, $p<.01$).

Generally, Clusters 5.1, 5.2, and 5.3 performed worse across several neuropsychiatric domains than Clusters 5.4 and 5.5 ($p<.05$, $p<.01$, $p<.001$; Table D4.3, Figure D4.3c). Yet, there was additional delineation among the three clusters with individuals with PTSD. For example, Cluster 5.3 performed better in Learning ($p<0.05$) than Cluster 5.1, whereas Cluster 5.2 performed better in Performance IQ than Cluster 5.1 ($p<0.05$; Table D4.3, Figure D4.3c). Cluster 5.2 had smaller average bilateral frontal volumes than Cluster 5.4 ($p<0.05$) and smaller average

bilateral temporal cortex volumes than Clusters 5.4 and 5.5 ($p<0.01$, $p<0.001$; Table D4.3, Figure D4.3d). Interestingly, Cluster 5.3 also had significantly larger average right temporal cortex volumes than Cluster 5.2 ($p<0.001$), while Cluster 5.2 had more right frontal white matter than Cluster 5.3 ($p<0.05$). Cluster 5.5 had larger average bilateral parietal, occipital, and right temporal CSF than Clusters 5.2 or 5.1 (Table D4.3, Figure D4.3d). Finally, Clusters 5.4 and 5.5 had higher average baseline Day 1 cortisol levels than Cluster 5.3, and greater cortisol suppression after DST than Cluster 5.1 (Table D4.2, Figure D4.3e).

Test	Cluster 1 Mean	S.D.	Cluster 5.3 Mean	S.D.	Cluster 5.2 Mean	S.D.	Cluster 4 Mean	S.D.	Cluster 5 Mean	S.D.	ANCOVA P-value	Sig	F	Post Hoc Cluster Difference TK
Neuropsychiatric (D.F.=4,725)														
Executive Functioning 1	8.61E+00	2.29E+00	9.17E+00	3.35E+00	9.76E+00	3.32E+00	1.10E+01	2.83E+00	1.08E+01	3.12E+00	1.83E-04	***	5.74	4>1**,4>2**,5>1*, 5>2*
Executive Functioning 2	1.07E+01	2.35E+00	1.05E+01	3.09E+00	1.07E+01	2.69E+00	1.22E+01	2.29E+00	1.17E+01	2.97E+00	1.10E-04	***	6.04	4>1*,4>2**,4>3*
Learning 2	1.04E+01	3.17E+00	1.09E+01	2.53E+00	1.16E+01	2.65E+00	1.26E+01	2.27E+00	1.23E+01	2.82E+00	7.94E-05	****	6.23	4>1**,4>2**,5>2*
Learning 3	4.17E+01	1.05E+01	4.56E+01	1.04E+01	4.88E+01	1.03E+01	5.18E+01	9.34E+00	5.06E+01	1.02E+01	1.35E-05	****	7.22	4>1***,4>2**,5>1**,3>1*
Memory 1	9.35E+00	2.71E+00	9.33E+00	3.80E+00	1.02E+01	3.45E+00	1.16E+01	2.70E+00	1.13E+01	3.19E+00	6.76E-05	****	6.33	4>1*,4>2***,5>2*
Memory 2	1.02E+01	2.44E+00	1.02E+01	3.51E+00	1.09E+01	3.03E+00	1.23E+01	2.55E+00	1.20E+01	2.86E+00	4.92E-05	****	6.51	4>1*,4>2**,5>2*
Performance IQ 1	5.58E+01	1.09E+01	6.60E+01	1.23E+01	6.50E+01	1.60E+01	6.70E+01	1.55E+01	7.00E+01	1.31E+01	3.05E-04	***	5.44	4>1**,3>1*,5>1***
Performance IQ 2	3.42E+01	1.02E+01	3.76E+01	1.25E+01	3.97E+01	1.20E+01	4.26E+01	1.16E+01	4.48E+01	1.17E+01	6.48E-05	****	6.35	4>1*,5>1**,5>2*
Verbal IQ 1	2.20E+01	5.72E+00	2.39E+01	3.95E+00	2.26E+01	4.93E+00	2.44E+01	4.77E+00	2.43E+01	4.81E+00	3.72E-02	*	2.59	N.S.
Verbal IQ 3	1.55E+01	4.02E+00	1.48E+01	3.44E+00	1.69E+01	4.14E+00	1.69E+01	4.35E+00	1.75E+01	4.14E+00	5.14E-03	**	3.78	4>2*,5>2**
Verbal IQ 4	1.72E+01	5.53E+00	1.90E+01	4.75E+00	1.91E+01	3.67E+00	2.02E+01	4.32E+00	1.98E+01	4.38E+00	2.19E-02	*	2.91	4>1*
Imaging Volumes (D.F.=4,162)														
Right Hippocampal	2.07E-03	2.00E-04	2.01E-03	3.00E-04	2.26E-03	3.00E-04	2.14E-03	3.00E-04	2.12E-03	2.00E-04	2.66E-02	*	2.83	3>2*
Right Lenticular White Matter	2.71E-03	8.19E-04	2.29E-03	4.41E-04	2.33E-03	4.62E-04	2.23E-03	4.65E-04	2.19E-03	4.24E-04	2.96E-02	*	2.76	D>4*,D>5*
Cerebellar White Matter	2.76E-02	2.85E-03	2.61E-02	3.51E-03	2.85E-02	6.30E-03	2.53E-02	4.78E-03	2.59E-02	4.59E-03	3.40E-02	*	2.67	D>1*
Right Parietal CSF	1.90E-02	4.30E-03	2.18E-02	4.70E-03	2.26E-02	4.70E-03	2.20E-02	5.00E-03	2.40E-02	5.20E-03	1.45E-02	*	3.2	5>1*
Left Parietal CSF	1.83E-02	4.30E-03	2.17E-02	4.60E-03	2.19E-02	4.80E-03	2.13E-02	4.40E-03	2.31E-02	5.50E-03	1.92E-02	*	3.03	5>1*
Left Occipital CSF	4.27E-03	1.20E-03	5.26E-03	1.20E-03	4.83E-03	1.20E-03	4.74E-03	1.20E-03	5.70E-03	1.50E-03	1.32E-04	***	6.11	5>1**
Left Temporal CSF	1.33E-02	3.19E-03	1.40E-02	2.55E-03	1.32E-02	2.39E-03	1.23E-02	2.32E-03	1.45E-02	3.24E-03	3.01E-02	*	2.75	N.S.
Subcortical CSF	8.95E-04	2.24E-04	1.05E-03	2.70E-04	1.03E-03	1.97E-04	1.03E-03	1.97E-04	1.08E-03	2.44E-04	2.58E-02	*	2.85	4>1*

D.F., degrees of freedom
S.D., standard deviation
TK, Tukey-Kramer
N.S., not significant
*, p<.05; **, p<.01; ***, p<.001; ****, p<.0001

Table D4.3. Biomarker Differences Across Clusters in the CS Dataset for k=5

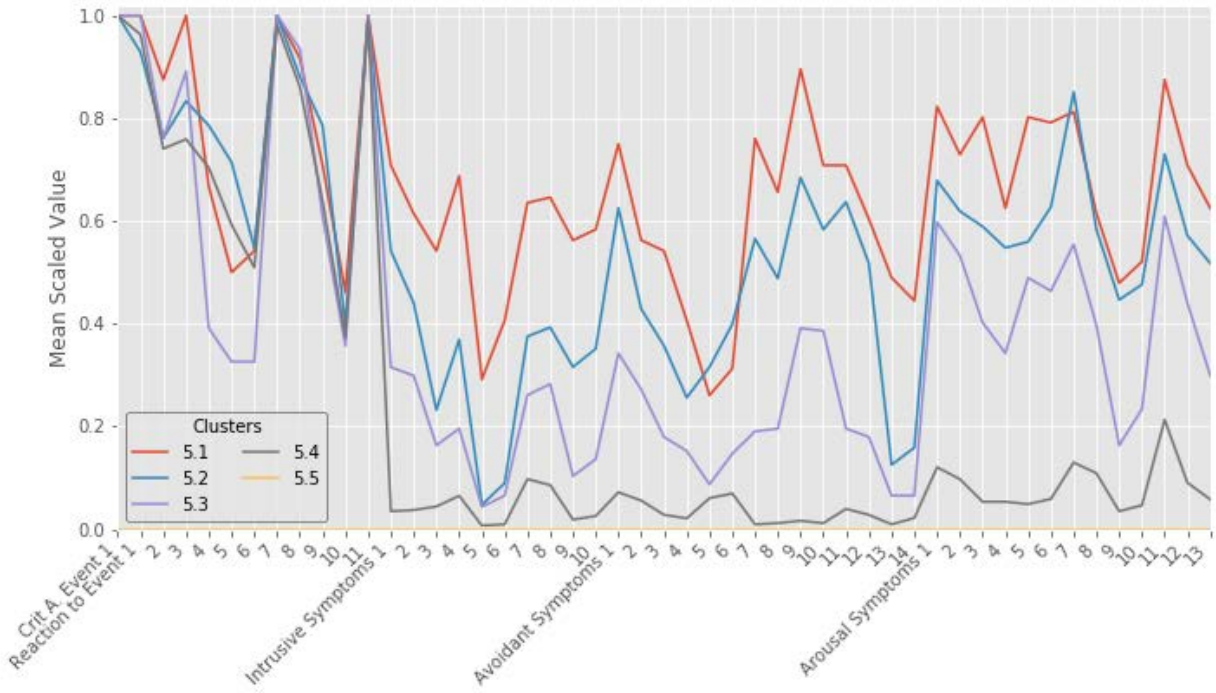

Figure D4.3a. Symptom Profiles for All Items in the Clinician Assessment for PTSD (k=5)

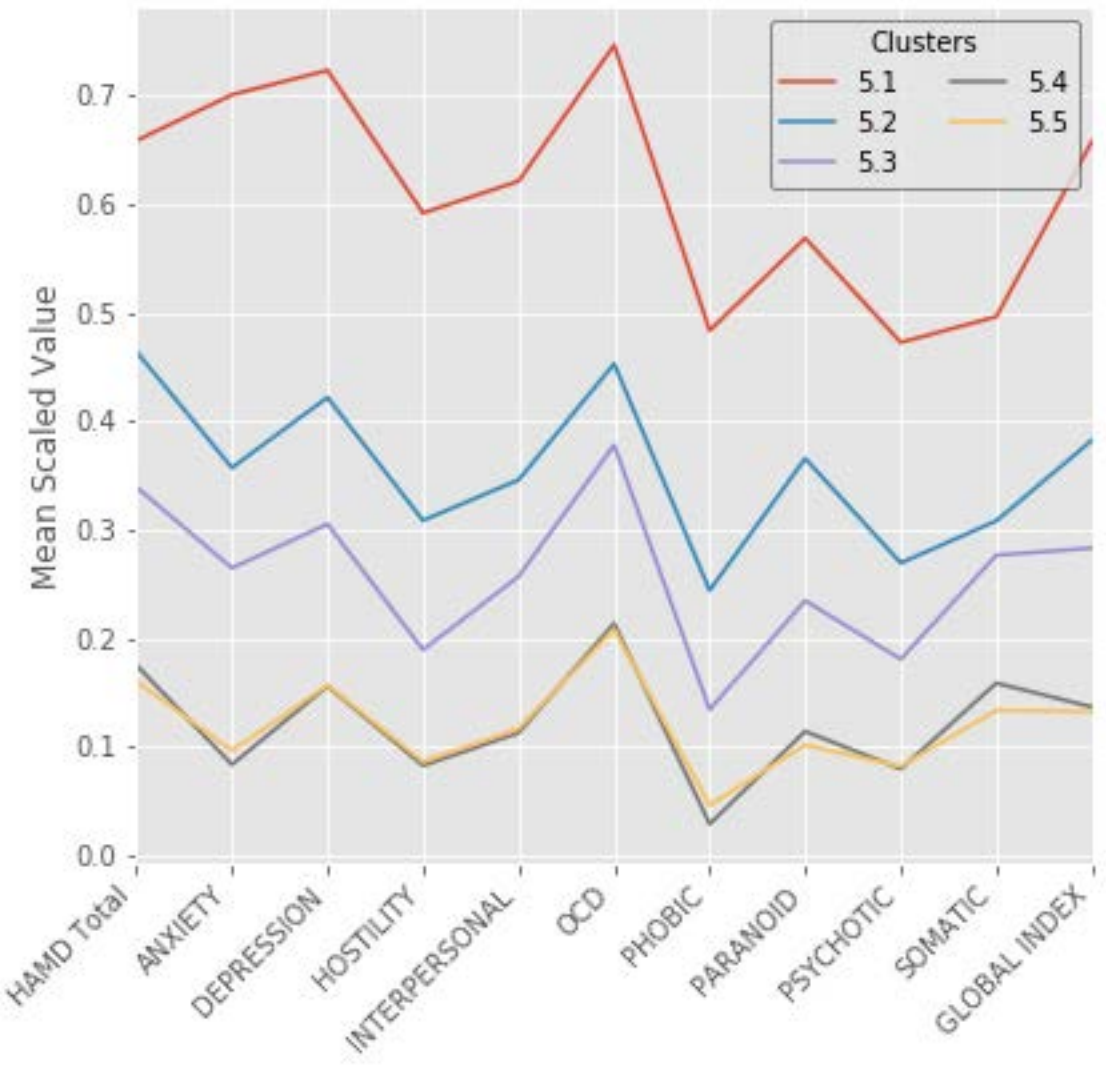

Figure D4.3b. Symptom Profiles for Summary Scores in Hamilton Depression Inventory and Symptom Checklist-90 (k=5)

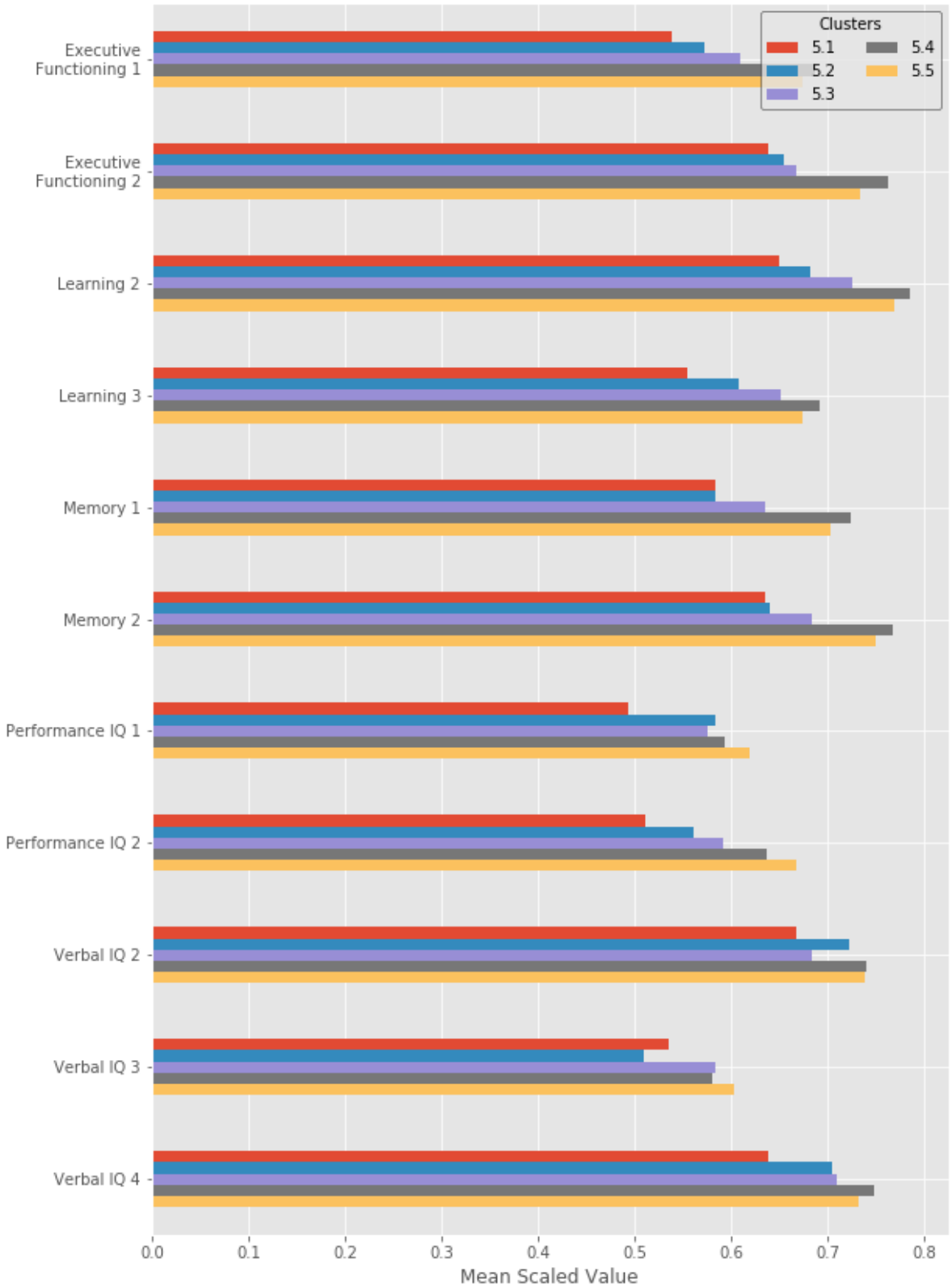

Figure D4.3c. Neuropsychiatric Markers with Significant Differences Across Clusters (k=5)

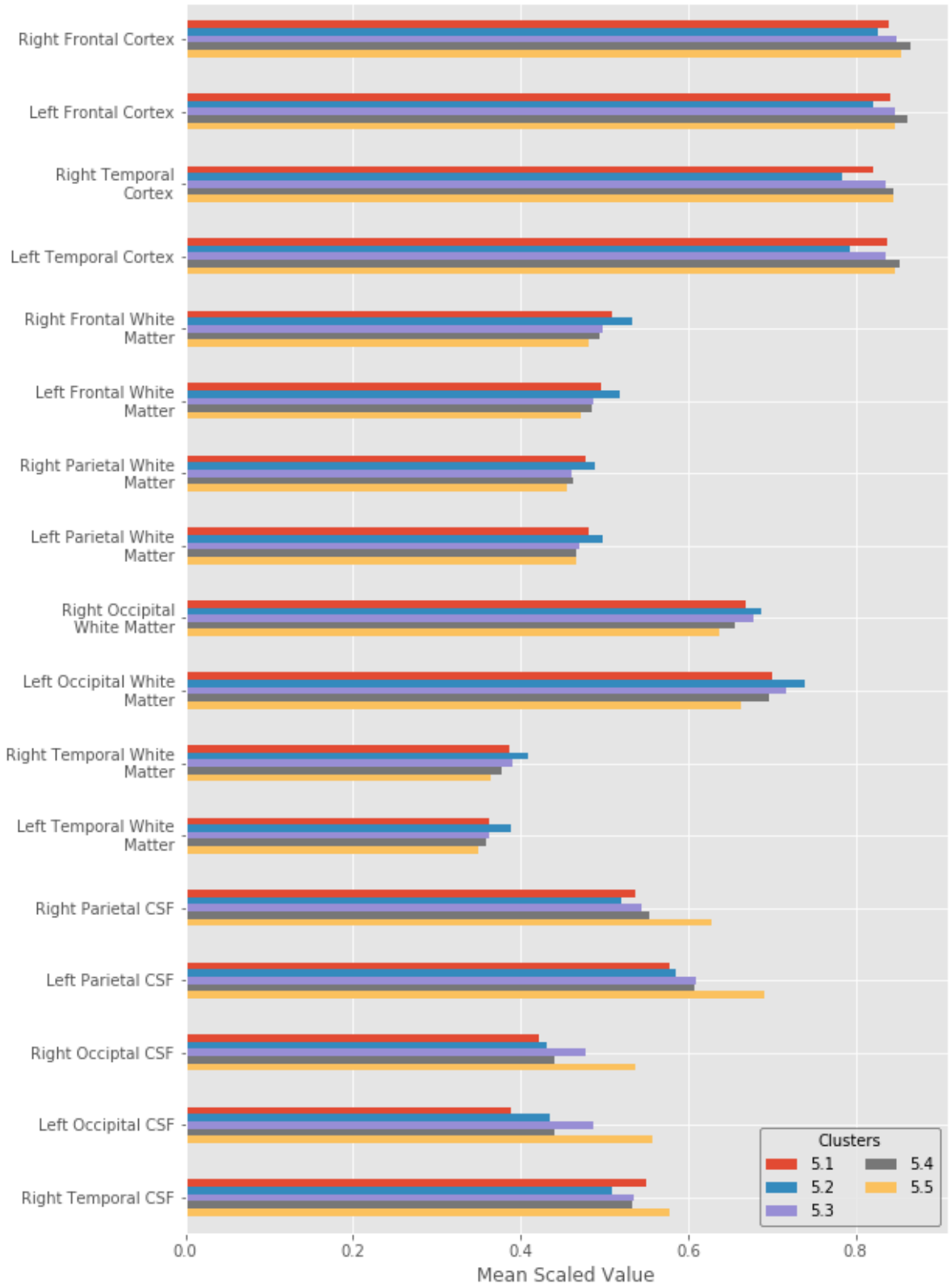

Figure D4.3d. Imaging Markers with Significant Differences Across Clusters (k=5)

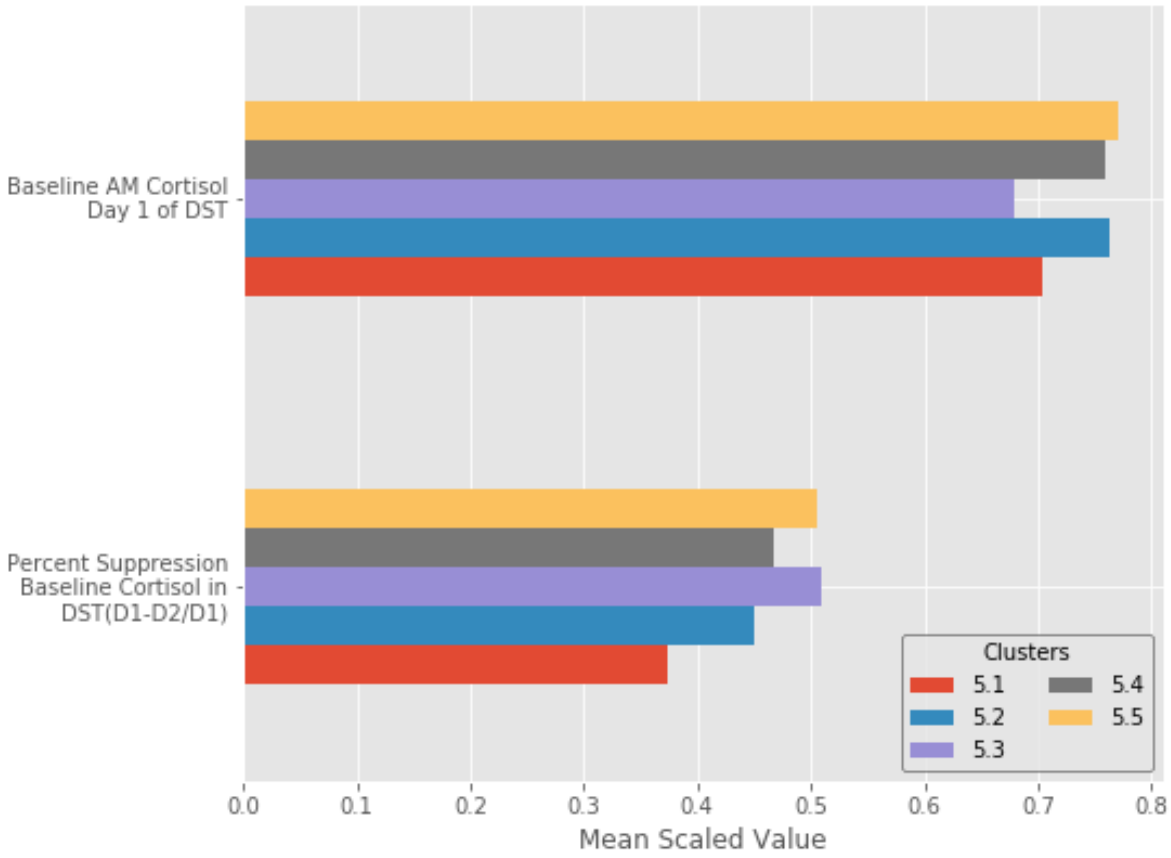

Figure D4.3e. Endocrine Markers with Significant Differences Across Clusters (k=5)

UCSF LIBRARY RELEASE FORM

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature  Date 03/29/2018