

**DEVELOPING A MANUALLY ANNOTATED CORPUS OF VA ELECTRONIC  
MEDICAL RECORD NOTES FOR POST-TRAUMATIC STRESS DISORDER  
NATURAL LANGUAGE PROCESSING TASKS**

**By**

**Maryan Zirkle**

**A THESIS**

**Presented to the Department of Bioinformatics and Epidemiology  
and the Oregon Health & Science University  
School of Medicine  
in partial fulfillment of  
the requirements for the degree of**

**Master of Science**

**September 2013**

## Certificate of Approval

This is to certify that the Master's Thesis of

**Maryan Zirkle**

*“Developing a Manually Annotated Corpus of VA Electronic Medical Record Notes for Post-traumatic Stress Disorder Natural Language Processing Tasks”*

Has been approved

\_\_\_\_\_Holly Jimison PhD\_\_\_\_\_

Thesis Advisor

\_\_\_\_\_Melissa Haendel PhD\_\_\_\_\_

Committee Member

\_\_\_\_\_Jessica Ross MD\_\_\_\_\_

Committee Member

\_\_\_\_\_

Committee Member

## TABLE OF CONTENTS

List of Tables.....	ii
List of Figures .....	iii
Acknowledgements .....	iv
Abstract .....	v
Chapter 1 Introduction.....	1
Chapter 2 Background.....	3
Chapter 3 Materials and Methods .....	7
Chapter 4 Results.....	17
Chapter 5 Discussion.....	22
Chapter 6 Summary and Conclusions .....	28
Chapter 7 References .....	30
Chapter 8 Tables.....	32
Chapter 9 Figures .....	38
Chapter 10 Appendix.....	42

## LIST OF TABLES

<i>Number</i>	<i>Page</i>
1. Overall Difference Between Batch 1 and Batch 2 .....	32
2. Batch Difference Based on Total Annotation for Specific Classes .....	32
3. Inter-Annotator Agreement (IAA) for Batch 1 and Batch 2.....	33
4. Average Inter-Annotator Agreement (IAA) for Batch and Class.....	33
5. Number of Total and Unique Annotations for Symptomatology Elements .....	34
6. Number of Total and Unique Annotations for Treatment Elements .....	34
7. Number of Most Frequently Annotated Concepts for Symptoms and Treatments.....	35
8. Number of Unique Annotations for Each Treatment Category .....	35
9. Example of Relationship Between Documented Symptom and Treatment Prescribed .....	36
10. Relationship Between Documented Symptom Concept and Treatment Category: Pharmacologic and Psychological.....	36
11. Relationship Between Documented Symptom Concept and Treatment Category: Psycho-Education and Psychological .....	37
12. Relationship Between Documented Symptom Concept and Treatment Category: Pharmacologic, Psycho-Education and Psychological.....	37

## LIST OF FIGURES

<i>Number</i>	<i>Page</i>
1. Sample Annotated Document .....	38
2. Evolution of Agreement for Symptom Pass.....	39
3. Evolution of Agreement for Treatment Pass .....	40
4. Number of Unique Annotations for Each Treatment Category .....	41
5. Venn Diagram: Relationship Between Documented Symptom Concept and Treatment Category Prescribed (Pharmacologic, Psycho-Education, and Psychological) .....	41

## **ACKNOWLEDGMENTS**

The author wishes to express sincere appreciation to Professors Haendel and Jimison for their assistance in the preparation of this manuscript. In addition, special thanks to Drs. David Hickam, Jessica Ross, and Dezon Finch whose familiarity with the needs and ideas of the study were helpful during this research effort. Thanks also to team members of the Department of Veterans Affairs Consortium for Healthcare Informatics (CHIR) for their valuable input.

## ABSTRACT

Mental health notes contain implicit and explicit concepts difficult to extract from the narrative note. This “free text” contains vital information necessary for comprehensive terminology creation used to enhance Natural Language Processing (NLP) tasks and is an ongoing challenge for use with advanced NLP tools. This project will build a gold standard corpus of narrative, clinical text manually annotated for key clinical data describing several concepts documented within the Veterans Affairs (VA) for Post-traumatic Stress Disorder (PTSD); most notably symptomatology and treatment modalities. We describe and discuss the annotation technique with regard to process and content, which includes defining schema, creating guidelines (determining the level of concepts to be annotated), and using a double annotation method. More than 900 clinical documents from the VA electronic medical record (EMR) for patients known to have a diagnosis of PTSD are used. Protégé 3.4.8-Knowtator was used for annotation and all clinical documents are analyzed and remain in the VA’s secure Informatics and Computing Infrastructure (VINCI). Annotator agreement is reported and each document is adjudicated by an expert clinician. We report general results on the final dataset created. We also report inter-annotator agreements (IAA), document batch differences, and provide descriptive statistics that query the data by type, frequency, and relationship to illustrate the main features of the resulting data. The annotated corpus of clinical notes will be used to facilitate NLP tasks in named entity recognition, automated categorization, and temporal analysis. The validated annotations and unique terms will also be used as a supporting resource of terminology and relationships for incorporation into an ontology.

## CHAPTER 1: INTRODUCTION

Veterans are a significantly large group among the population suffering from mental health disorders. In 2009, it was reported that 37% of veterans returning from Iraq and Afghanistan have mental health problems.<sup>1</sup> A study done by Seal et al on the prevalence of mental health disorders among Operation Enduring Freedom and Operations Iraqi Freedom (OEF/OIF) veterans from 2002-2008 showed a six fold increase in new mental health diagnoses from the beginning of the study in 2002. More specifically, PTSD constitutes a substantial proportion of the burden of illness among veterans.<sup>2</sup> A National Vietnam Veterans Readjustment Survey reported that 30.9% of all men who served in Vietnam developed PTSD. Surveys of military personnel returning from OEF/OIF have shown that 12.6 % of U.S. men who fought in Iraq developed PTSD and 6.2% of U.S. men who fought in Afghanistan developed PTSD.<sup>2</sup> In a study of 103,788 OEF/OIF veterans seen at VA health care facilities between September 2001 and December 2005, PTSD was the most commonly diagnosed military service-related mental health diagnosis (13,205 cases), accounting for more than half of the veterans receiving a mental health diagnosis and 13% of all OEF/OIF veterans in the study.<sup>3</sup> During Fiscal Year (FY) 2006, VA medical center programs served over 346,000 veterans diagnosed with PTSD in specialized outpatient programs and general mental health clinics.<sup>4</sup>

The obscurities of PTSD have been subject to considerable debate and controversy over the years.<sup>5</sup> PTSD is heterogeneous with regard to the expression of symptoms, the severity of symptoms, and the chronicity of symptoms.<sup>2</sup> The course of this disorder may vary as well, in



the duration of symptoms and the level of difficulty in functioning.<sup>6</sup> It has also been shown that there may be some unique aspects of PTSD encountered by veterans versus civilians.<sup>2</sup> Moreover, when dealing with treatment of this disorder, there was a study by the Department of Defense and Department of Veterans Affairs that determined serious inconsistencies in research on treatment modalities which led to a difficulty understanding if the current treatment interventions were useful to help treat this disorder.<sup>2</sup>

## CHAPTER 2: BACKGROUND

The goal of this study is to create a gold standard corpus of clinical text manually annotated for language describing the complex concepts of symptoms and treatments for patients diagnosed with PTSD. Many advanced NLP systems rely on the information contained in these clinical documents, yet the implicit and explicit concepts can be difficult to extract from narrative text. The detailed concepts need to first be manually annotated for quality gold standard creation and then with directed evaluation of NLP tools we can eventually answer some of the looming questions embedded in these controversies and inconsistencies. First, defining the vocabulary used by clinicians can assist automated methods for extracting clinical details from the text contained in notes.<sup>7</sup> Here we describe the annotation of clinical notes from the VA EHR, VistA. We annotate symptom, symptom change, treatment, treatment change, treatment non-compliance, negation, and explicit instances of symptom targeted by treatment. Intended application of this corpus includes, but is not limited to automated detection of similar concepts on larger corpora, automated categorization of psychotherapeutic and pharmacologic treatments and techniques, and further temporal concept annotation using the created gold standard dataset. The terminology extracted will also be used to inform the creation of PTSD ontology to enhance NLP tool creation.

In the medical field clinical notes serve several functions. From the beginning, these notes served as documentation of the patient-provider encounter and a method of communication between providers. Since the advancement of the electronic medical record (EMR), these documents have also served as a systematic collection of health information about patients and

populations, capturing copious amounts of text on individual clinical encounters. The problem is much of the documentation for these medical encounters is created in “free text” form. This means the documentation is unstructured in that it is not embedded in a template wherein the context is specified. Within the department of Veterans Affairs, much of the clinical note is in mixed template and free-text form, creating an even more unpredictable format that is often difficult to utilize.<sup>8</sup> Physician documentation contains elements that would pose a challenge to an NLP tool process; for example, medical language abbreviations and negated concepts can be embedded in the free text of a document.<sup>9</sup>

The successful extraction of meaning from the unstructured documentation would serve to create improved clinical data analysis. Natural language processing (NLP) tools have been designed to extract this data and give this narrative, free form body of text more meaning. In order to develop these NLP tools, a gold standard of robust, manually annotated materials is typically required. Obtaining these gold standard annotations for use in the development of NLP tools is difficult due to the time and labor involved.<sup>10</sup> Manual annotation relies on standardized guidelines to determine what is incorporated. They must be specialized and are often domain specific.<sup>9,11</sup> These guidelines are developed through an iterative process, which is readily accessible and discussed among annotators after a consensus mode and expert adjudication.<sup>12</sup> Though the annotation process attempts to strike a balance between achieving maximum accuracy, speed, and agreement, there is often a need to avoid the prohibitive expense of clinical expertise, whose time may be difficult to obtain.<sup>13</sup> Some have used pre-annotated corpora to test their NLP tools,<sup>14</sup> while others have placed annotation focus on more well defined concepts such as protected health information, which has been shown to be

easier to detect and was associated with increased annotator accuracy.<sup>15-16</sup> Others rely entirely on expert clinician annotations.<sup>17-18</sup> Some teams found non-clinicians achieved agreement with clinician annotators, but more training and time was involved.<sup>19</sup> For this study, we chose to annotate for concepts that had not yet been defined. This made pre-annotation techniques and the use of previously well-defined guidelines virtually impossible. Therefore, expert clinicians were used to evaluate and adjudicate the final annotations.

In the recent literature, annotation has been the standard technique used to assess the performance of the extraction of clinical concepts from EMR text in several clinical domains: indicators of heart failure in clinical notes for primary care patients,<sup>10</sup> pathology reports,<sup>15,20-21</sup> Emergency Department reports,<sup>19</sup> radiology reports,<sup>16,18-22</sup> smoking cessation guidelines,<sup>23</sup> medications,<sup>24,25</sup> chronic disease,<sup>8</sup> and disease outbreak reports.<sup>26</sup> However, there is a dearth of mental health based annotation efforts, and none that especially focus on PTSD. We suspect this is due to the difficulty in detecting these types of undefined concepts. There has been some work done on emotion detection and classification of human annotated suicide notes,<sup>27</sup> where “vested volunteers” were recruited to annotate concepts of emotion,<sup>28</sup> but to our knowledge, our study is the first study that has completed manual annotation on PTSD in mental health procedure notes. We set out to define the concepts through specific annotation guidelines (see Appendix) and expert consultation to better understand the language used to describe the clinical course of patients diagnosed with PTSD in the VA.

## GOALS

Create a gold standard corpus of clinical text for PTSD:

- AIM 1: Manually annotate for the complex concepts of symptoms and treatments.

- AIM 2: Develop an efficient annotation strategy.
- AIM 3: Query the annotated data by type, frequency, and relationship.
- AIM 4: Use validated annotations for the development of NLP tools for name entity recognition and the development of tools that use the terms in ontology.

## CHAPTER 3: MATERIALS AND METHODS

### Annotation Task

The task consists of annotating concepts related to symptoms and treatments of PTSD from clinical notes. The objective is to create a corpus of annotated text that identifies the language used by providers to document information about the symptoms suffered and treatments prescribed for patients diagnosed with PTSD. These efforts were separated into two passes (Figure 1). Pass 1 involves annotation of clinical notes with a specific schema to capture the varied manifestation of PTSD symptomatology. These symptoms are then post processed by expert clinicians into DSMIV key symptom categories. Pass 2 involves annotation of the same clinical notes with a specific schema to capture the prescribed psychotherapeutic and pharmacologic treatments for PTSD in the VA. These treatments are then categorized by the annotator and adjudicated by the clinician expert. These tasks were divided into two separate passes made on the same sets of clinical notes in order to limit the cognitive load on annotators.

### Corpora

All data was captured, analyzed and remains in the VINCI data warehouse. The documents contain highly sensitive patient information. Major efforts were employed to keep the information secure while conducting the research. Documents were selected across one fiscal year based on the ICD-9 code (309.81) that indicated a diagnosis of PTSD. These patients were also noted to have at least two outpatient visits for PTSD management between 10/1/2010 and 9/30/2011. We selected notes based on the Veterans Health Administration

(VHA) Enterprise Standard Title having the phrase “Mental Health” in the title and then selected 16 titles (N=1416008). We then checked to make sure the notes occurred at several different VA sites (119 sites). Notes were then selected using local titles such as: “Ind Therapy Note”, “CPT Ind Psychotherapy”, and “Procedure(s)”.

*Batch 1 Notes:*

A random sample was drawn from these notes to create a first batch of notes containing two sets of approximately 200 notes each to support Pass 2 and four annotation sets with approximately 100 notes each to support Pass 1. These were the same documents simply separated into 2 sets for one pass and 4 sets for the other pass. Initially, it was thought that completing Pass 2 first might help to orient the annotators to their task, but based on comments from the annotators, Pass 2 was more heavily laden with clinical jargon that was very domain specific. When text was relevant, it was often an extremely specific psychotherapeutic modality that was initially difficult for the annotator to grasp.

*Batch 2 Notes:*

In order to decrease the number of notes in the corpora that contained no relevant concepts for annotation and optimize our annotators’ time, a second batch was oversampled for documents identified as specific mental health treatment procedure notes. This was done where a count of annotations for treatments and symptoms for each note type from the first batch was used to narrow down and select the top four note types that contained the highest frequency of annotated text to support the task. This allowed us to then evaluate notes that were proven to have increased richness of PTSD concept documentation. This second batch

was separated into four annotation sets containing approximately 100 notes each for both Pass 1 and Pass 2. Pass 1 (symptomatology) was the initiating pass in this batch at the request of the annotators as it seemed like an easier initial task when compared to Pass 2 (treatment). Also, both passes were separated into four sets at the request of the annotators as more than 100 notes in a set were overwhelming.

### Annotation Guidelines

We developed our own guidelines for annotations of Pass 1 and Pass 2 (see Appendix).

#### ***Pass 1. Symptom elements:***

We define two classes to capture specific text related to symptomatology of PTSD:

- Symptom
- Symptom Change

The symptom class captured common PTSD symptoms listed in DSMIV diagnostic criteria for PTSD, any synonyms of those symptoms and any documentation of symptoms that were explicitly related to PTSD, but not defined in DSMIV. Common PTSD symptoms include terms such as: *avoidance, hallucination, flashback* and *re-experiencing trauma events*. Other terms such as: *shame, guilt, depressive symptoms*, and *fear* are examples of terms explicitly related to PTSD, but not defined in DSMIV. Annotator guidelines required the annotation of only those symptoms that were explicitly documented as being reported as symptomatology in the therapeutic encounter for PTSD in order to attempt to eliminate capturing symptomatology that may not be relevant to the task. For instance, a document may contain language referring to sleep problems from obstructive sleep apnea. This does not meet the criteria for inclusion into the



dataset as it is not a sleep disturbance explicitly documented as being reported for symptomatology in a therapeutic encounter for PTSD. The symptom change class captured the relationship of a symptom and its surrounding modifier or change language. This class identified any documentation by the provider about an observed change in symptoms or patient reported worsening or improving symptoms. For example, if the document contained the text, “increased nightmares” the annotator would capture *nightmares* as a symptom and *increased* as a symptom change linked to this specific symptom.

Two other classes were also used to support relationships among the concepts within the free text:

- Negation
- Experiencer

The negation class captured any term used to negate a documented PTSD symptom. This required the annotation of both the symptom and the negation term so that a relationship was created to convey that the specific concept was negated. For example, if the document contained the text, “denies suicidal thoughts” the annotator would designate *suicidal thoughts* as an instance of the symptom class and *denies* as an instance of the negation class linked to this specific symptom. The experiencer class captured the individual experiencing the symptom annotated. This was typically captured as *he/she* or *Veteran*. This also included the option to designate if this was the patient or someone else (i.e. patient’s wife, another patient in a group setting).

***Pass 2. Treatment elements:***

- Treatment (psychotherapeutic and pharmacologic modalities)
  - Targeted symptom
- Treatment Change
- Treatment Non-compliance
- Dosage
- Duration
- Frequency
- Provider
- Audience

The treatment class was divided into six categories: psychological, psychosocial, psycho-education, case management, pharmacologic, and other. These categories were designated by the annotator as a slot value relationship linked to the treatment annotated. The categories include:

- Psychological

Any mention of regular sessions with mental health providers to help patients understand why they are acting and thinking in ways that are troubling or dangerous to themselves or others so they have more control over their behaviors and can correct them. This includes therapies such as: Cognitive Behavioral Therapy, Cognitive Processing Therapy, Dialectical Behavioral Therapy, and Trauma Focused Therapy. This can also include general therapy techniques such as building a rapport or active listening.

However, this will not include typical generic communication language such as praise or encouragement that is not tied to a specific therapeutic intervention.

- Psychosocial

Any combination of psychotherapy and social or communication training.

This includes mention of recreational and social interaction techniques.

- Psycho-education

Any mention of teaching patients about their illness, how to treat it, and how to recognize signs of relapse so they can get necessary treatment before their illness worsens or occurs again. This includes teaching coping strategies and problem-solving skills. This excludes any educational Cognitive Behavioral Therapy techniques that should fall under the psychological category.

- Case Management

Any collaborative process dealing with assessment, planning, facilitation and advocacy for options and services to meet a patient's all-inclusive needs through communication and available resources to promote quality, cost-effective outcomes. This will include any discussion of treatment goals or treatment planning.

- Pharmacologic

Any medication prescribed or any over the counter medication currently taken within the confines of the therapeutic relationship for the treatment of

symptoms (i.e. depression, anxiety, sleep disorders) specifically related to PTSD.

- Other

Any text related to treatment for PTSD that does not fulfill the requirements of the other five categories. This may include chapel services, admission to the hospital, and the use of restraints.

The symptom targeted by the documented treatment was captured as a slot value of the treatment class when available for annotation. This included any explicit documentation of a treatment and its targeted symptom. The treatment change class captured the relationship of a treatment and its surrounding modifier or change language similar to the method described for the symptom change class. This class included any documentation by the provider about changes in treatment. This did not include any hypothetical changes in treatment. The treatment non-compliance class captured any documentation that a patient did not adhere to prescribed treatment protocol/guidelines. This included documentation of terms/phrases such as: patient no-showed for psychotherapy treatment or the patient stopped taking a prescribed pharmacologic treatment. The treatment dosage class captured the dose of medication prescribed. This was typically documented in milligram form. The treatment duration class captured the length of a psychotherapy session or the duration of time for a specific prescribed medication. The treatment frequency class captured how often a patient would take a certain medication or in some cases the number of times a certain task was to be completed in a therapy session. The treatment provider class captured the credentials and possibly the title of the provider engaged in the documented encounter. The treatment audience class captured the

setting in which the documented encounter took place. This would include the designation of whether an encounter involved an individual, couple, or group. A treatment negation class was not necessary because annotations were only to be made of those treatments that were given during the encounter.

### Annotators

Two annotators performed the annotation tasks. Two clinicians annotated the first batch of notes. Two non-clinicians (with Masters Degrees in a non-mental health discipline) annotated the second batch of notes.

### Software

The Protégé 3.4.8 (<http://protege.stanford.edu>) plug-in Knowtator (<http://bionlp.sourceforge.net/knowtator>) software was used for the annotation of documents in our corpora. All of the accessed software and documents were stored on our Veterans Affairs Informatics and Computing Infrastructure (VINCI) in compliance with HIPAA and IRB regulations. VINCI can be accessed from within the VA and remotely via CITRIX VPN under secure connections. All annotated data remains in VINCI for use in future research on veterans with PTSD.

### Annotation Process

Two annotators annotated each document. This is referred to as double annotation. The annotations of each annotator are converted into a consensus set whereby expert adjudication is performed and a final annotated document is created. Annotators were trained on technical

aspects of the software tool as well as the target clinical concepts for each pass. Training notes were annotated before the official task in order to evaluate the annotator's grasp of the task. Annotators did not begin the official annotation on either task until satisfactory agreement was reached during the training task. After double annotation of each document, a clinician adjudicated the annotations made by the two annotators for each set of notes and discrepancies were addressed and incorporated into annotation guidelines with appropriate examples where necessary. The annotators were made aware of any changes/additions to the guidelines upon completion of each set of notes and before beginning the annotation of the subsequent set of notes. The expert clinician adjudicators evaluated the annotations in a consensus mode first. It is important to note that in this consensus mode the expert sees only those annotations that were not agreed upon by the two annotators. The expert accepts what is correct thereby deleting any incorrect annotations. In some cases, the expert will delete both and create an annotation for the correct concept. In order to prevent errors in a circumstance where both annotators miss a concept and thus it never appears in the consensus mode for adjudication, the adjudicator read each document for any such "double misses" by the annotators. This adds significantly to the integrity of the dataset.

#### Inter-Annotator Agreement Measurements

Typically, F-measures are reported for named entity annotation tasks in text documents because there is no predetermined number of items to contend with when annotating text that is to say, there is no known comparison.<sup>16</sup> Cohen's Kappa scores are typically reported for classification annotation tasks because it requires the number of negative cases to be computed. As mentioned, we do not know for certain the number of negative cases in a

named entity annotation task.<sup>16</sup> However, we do report annotator agreement where the annotations of one annotator are compared to the annotations of the other annotator in order to ensure that there is suitable agreement of concept extraction in the task. The use of the expert clinician as adjudicator and the additional evaluation of any possible “double misses” in the annotation task are suitable for saying this is a gold standard dataset without reporting specific F-measures.<sup>16,29</sup>

## CHAPTER 4: RESULTS

There were a total of 911 annotated notes. Out of the 911 notes, the largest amount of annotations in a single note was 84. 517 notes had five or more annotations per note. 402 notes had 10 or more annotations per note. There were a total of 612 patients from 119 different VA sites throughout the country. Out of 612 patients, the largest amount of notes from one patient was 18. 35 patients had five or more notes. Out of 119 sites, the largest amount of notes from a single site was 179.

### Batch Differences

Using two batches of notes allowed for the opportunity to examine the typical issues that come into play when undertaking this sort of an annotation project. There were only 2012 total annotations in the first batch of 585 clinical notes. Yet, there were 2941 total annotations in the second batch of only 326 clinical notes. Total annotations include instances of repeated concepts. There were 572 patients represented in the first batch of clinical notes, but only 44 of the total 572 patients represented in the second batch of clinical notes. Finally, there were 119 VA sites represented in the first batch of clinical notes with only 6 sites of the total 119 VA sites represented in the second batch of clinical notes. Refer to Table 1 for a summation of this data.

These results are evidence that an oversampled document selection process captures increased richness in the language used to describe the targeted complex concepts.



As shown in Table 2, there were 782 total annotations captured for the class of symptom in the clinical notes for the first batch of documents. There were 1528 total annotations for this same class in the second batch of documents. This resulted in almost double the amount of total annotations. The class of symptom change also contained almost two times as many total annotations in the second batch of clinical notes when compared to the first batch (Table 2). The class of negation contained more than four times the amount of annotations in the second batch as contained in the first batch of clinical notes: 249, 54 respectively (Table 2). Conversely, there were 447 total annotations for the class of treatment in the clinical notes for the first batch of documents, but only 501 total annotations (54 more annotations) for this same class in the second batch of documents. This seems odd due to the otherwise increased richness of Batch 2, but it is important to remember that this is 501 annotations from only 326 clinical notes.

### Inter-Annotator Agreement

Table 3 shows the inter-annotator agreement (IAA) numbers for each set of notes annotated in each batch of notes. The IAA average is based on the comparison of the annotations from each annotator (Table 4). After the annotators complete a set of notes, their work is compared during an adjudication session where a consensus is determined. The rules for annotation of the text are then articulated and incorporated into the guidelines. The goal is to exhaust the introduction of new rules into the guidelines so that agreement tends to level off. This did eventually happen for both passes, but was most pronounced in the symptomatology annotation pass (Figures 2 and 3).

## Descriptive Statistics (Type, Frequency, Relationship)

Here we report information about the annotated dataset based on type, frequency, and relationship.

### Type:

The symptom class had the largest number of total annotations overall. The treatment change category has the least number of annotations overall (Tables 5 and 6). The amount of unique terms is of greatest use for continued work in natural language processing and ontology creation. We extracted more than 1000 unique symptom concept terms and more than 800 unique treatment concept terms (Tables 5 and 6). Unique annotations include the distinctive concepts that were annotated without taking repeated instances into account. Below are examples of actual text and corresponding categorization extracted from the clinical document for symptoms and treatment, respectively.

## **PTSD Symptom Categories and Sample Terms**

### **ClusterB**

**Dreams** - *NMs*  
**Flashbacks** - *Hearing explosions*  
**Physical Reactivity** - *Shaking*  
**Psychological Distress** - *Catastrophizing*  
**Recollection** - *Intrusive memories*

### **ClusterC**

**Avoid** - *Isolates*  
**Detachment** - *Numbness*  
**Diminished Interest** - *Low motivation*  
**Foreshorten Future** - *Future cut short*  
**Poor Recall** - *Fragmented memory*  
**Restrict Affect** - *Blunted*

### **ClusterD**

**Difficult Concentrate** - *Easily Distracted*  
**Difficult Sleeping** - *Wakes up*  
**Hyper-vigilance** - *Watchful*  
**Irritability/Anger** - *Tense*  
**Startle Response** - *Jumpy*

### **No Cluster**

**General** - *Embarrassed*

*Italics= actual text*

23

## PTSD Treatment Categories and Sample Terms

<b>Case Management</b>	<b>Psychosocial</b>
<i>Connection claim</i>	<i>Social contact</i>
<i>case management needs</i>	<i>Communication</i>
<i>encouraged to speak to psychiatrist</i>	<i>Recreation</i>
<b>Pharmacologic</b>	<b>Psychological</b>
<i>Ambien</i>	<i>ABC sheets</i>
<i>Wellbutrin rs</i>	<i>Active listening</i>
<i>Diphenhydramine</i>	<i>Challenging beliefs</i>
<b>Psycho-education</b>	<b>Other</b>
<i>Anger management</i>	<i>Admission</i>
<i>Changing patterns</i>	<i>Restraints</i>
<i>Communication skills</i>	<i>Chapel services</i>

*Italics= actual text*

33

### Frequency:

With respect to frequencies of specific symptom concepts, the concepts of *suicide* and *homicide* were some of the most commonly represented concepts in the dataset (Table 7). Concepts such as *anxiety*, *anger*, *depression*, and *avoidance* were also repeatedly represented (Table 7). With respect to frequencies of specific treatment concepts, the concepts of *therapy*, *psychotherapy* and *cognitive therapy* were the most commonly represented concepts in the dataset (Table 7). The representation of these general treatment terms is expected based on the target domain. However, we did not anticipate the frequency of annotation to be so extraordinary for the concepts of *identifying stuck points*, *trauma account*, and *worksheet (homework)* (Table 7). Providers often document information pertaining to attempting to identify stuck points in a patient's thinking patterns, dealing with writing or reading trauma accounts, and assigning or reviewing homework in the form of worksheets.

Of the six treatment categories, the slot value for *Psychological Treatment* was most frequently used. This category involved all mentions of Cognitive Behavior Therapy which is a first line psychotherapeutic treatment used for patients with PTSD (Table 8). When looking at Figure 4, it is apparent that this is more than 50% of the total unique annotations for the treatment categories. We find that *psychological*, *psycho-educational* and *pharmacologic* are the most frequently documented treatment categories of the six possibilities available.

*Relationship:*

Results report 788 instances of relationships between documented symptom and treatment provided. Examples can be seen in Table 9. When looking at documented symptoms rolled up into a parent concept compared with the relationship of symptom targeted we find common patterns we can report for these relationships. For instance, Tables 10, 11, and 12 show the combination of Pharmacologic/Psychological, Psychological/Psycho-Education, and Pharmacologic/Psycho-Education/Psychological. You can see from these tables how each of the combinations of treatments is documented as being used to treat certain symptom concepts. Of note is the relationship between three of the most frequently documented treatment category types and the targeted symptom captured as “PTSD Symptoms”.

## CHAPTER 5: DISCUSSION

### Batch Differences

Ultimately, the sampling technique served to further the understanding of the intricacies involved in developing methods of solid, reproducible annotation in one of the most challenging clinical domains. First, oversampling for notes that were known to contain targeted concepts did contribute more information about the language used to describe the clinical course of symptomatology as it relates to PTSD. It did also provide a sense of longevity with representation of fewer patients having more notes over a longer period of time. However, it did not necessarily prove to increase the richness of data collected for the concepts related to treatment. It is important to note this may be due to the single annotation of multiple instances of these treatments in the treatment pass. This overall difference in the amount of total annotations as compared with data collected for the concepts related to symptoms could also be due simply to diligent documentation of symptoms by providers in the second batch of oversampled notes. Specific treatment encounters were the major identifying component for inclusion into the second batch of notes. It makes sense that providers would be more likely to document any and all symptoms reported in a specific treatment encounter as opposed to any other type of general encounter.

### Inter-Annotator Agreement

As can be seen best in Figures 2 and 3, the evolution of inter-annotator agreement for symptoms and treatments differs on several levels. For symptom annotation there is an increase from set 1 to set 2 that seems logical due to an expected learning curve for the

material. However, there is a marked decline in set 3 that can only be attributed to the introduction of new information into the guidelines (Table 3). For the remainder of the sets, the agreement once again improved and then eventually leveled off when the available information for annotation had been exhausted and no new information was being captured in the notes. For treatment annotation, there is a more consistent level of annotator agreement, although we do see an unexpected decline in set 3 (Table 3). This can also be attributed to the introduction of new information with a need for guideline clarification. For example, annotator agreement is calculated on class level (treatment type) as well as category level (treatment category). This means that if each annotator captured the correct concept at the class level, but disagreed at the category level, then there would be non-agreement and vice versa. This specific decline appears in two consecutive sets and is an indication of a weakness in the guideline coding definition for categorization (Figure 3). This is then corrected for in the subsequent set as seen by a rise in IAA scores in Table 3. Of note, are the higher IAA scores overall for batch two (Table 3). It is believed that the richness in the content of the notes in batch two allowed for the annotators to hone their content skills faster. Another area of interest when comparing the two batches is the use of different annotators. As previously stated, the first batch of notes was annotated by clinicians and the second batch of notes was annotated by non-clinicians. Studies have shown that non-clinicians are as accurate as clinicians when it comes to annotating clinical concepts given more time. Although, it is also reported that this time must be used with consecutive annotation requiring minimal breaks in the learning process and the overall task.<sup>17-19</sup> In our case, too many variables come into play to isolate any one reason for our increased extraction and agreement. Most likely a combination of more detailed guidelines and richer data with minimal interruption in the annotation process

has led to an efficient and comprehensive extraction of the key concepts from our designated dataset.

### Descriptive Statistics (Type, Frequency, Relationship)

#### Type:

Of the 4051 total annotations for the symptom class, 1111 are unique (non-duplicate) annotations. Most of these unique terms can be considered synonyms of the terms used to describe the diagnostic criteria for PTSD. These unique terms will be mapped to the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) for possible inclusion into the language used in this semantic nomenclature. Of the 1706 total annotations for the treatment class, 818 are unique (non-duplicate) annotations. These unique terms have been to some extent mapped to SNOMED-CT and already more than half have been shown to not exist in the semantic nomenclature used to date. Figure 4 shows that the majority of information captured relative to treatment documentation had to do with the basic language used to describe general psychological psychotherapeutic modalities. Again, these consisted of instances of the documentation of terms like, *Cognitive Behavioral Therapy* and *Cognitive Processing Therapy* and the language used to describe the various techniques included in these types of therapy encounters.

#### Frequency:

Of note, with regard to the most frequently annotated symptoms and treatments, found in table 7, is the increased documentation of the concepts of *suicide* and *homicide*. It is important to note that these annotations were not necessarily asserted instances of these concepts. For

example, it is a requirement of the provider in the therapeutic encounter for mental health to assess an individual's suicidal and homicidal state and document that this has taken place. Therefore, most instances of these concepts were related to a documentation of the instance of "no" suicidal ideation/intent/plan and "no" homicidal ideation/intent/plan. This is why it becomes important in natural language processing to capture the language used and patterns used to document that an occurrence of the instance of a concept has *not* (negation) taken place. This will eventually aid in computer systems understanding this language as humans understand it. A discussion on the most frequently annotated concepts for treatment class is also central to a discussion about how the results of this study are vital to the future of NLP in this clinical domain. As stated in the results, it is typical to find concepts such as *Therapy* and *Cognitive Therapies* habitually annotated throughout clinical documents in this domain. However, the language used to describe the concepts of the identification of *stuck point(s)*, *the trauma account* and *worksheet(s)* were not foreseen. The knowledge of the use of this unusual, subjective documentation style during treatment encounters is imperative for effective and efficient natural language processing tools that attempt to identify treatment documented in the EMR for PTSD.

*Relationship:*

Table 9 reports some examples of relationships between documented symptoms that are targeted by treatment prescribed. These sorts of relationships and their surrounding language will help to support future information extraction of similar relationships for further understanding of what treatments are prescribed for specific symptoms in the VA for the PTSD population. Having the ability to identify the documented connection between



treatment and targeted symptom will increase our capability to effectively evaluate the treatment given to Veterans with PTSD. It brings us closer to answering the question: What treatment works best for a specific symptom?

#### *Limitations to the dataset*

The differences described between the batches of selected notes can be seen as a limitation. The second batch had much richer notes; however, it also contained several instances of cutting and pasting or personal provider templates that made for repetition of several strings of text across several notes. On the other hand, the first batch of notes contained many notes that would be considered outliers because they had nothing to do with our intended task and thus received no annotations. Mostly, these notes were diverse without repetitive strings of text, but were not as plush as the second batch.

#### *Additional limitations to the study*

Another, but most likely inevitable limitation to the study is the difficulty in separating out the many overlapping or similar symptoms that presented in these patients with existing co-morbid mental health disorders. Nearly 80% of patients diagnosed with PTSD also have another mental health disorder.<sup>1</sup> Although we attempted to alleviate this problem with the detail of our guidelines, it is almost impossible, even for clinicians, to determine if, in the documentation of a single encounter, the symptom is attributed to their PTSD or a conflicting, overlapping disorder. In the future, including a class for co-morbidities associated with PTSD into the schema would help to clarify specific instances where these similar, overlapping symptoms most often occur. Having more insight into this aspect of PTSD symptomatology

would improve our ability to extract the most relevant data for use in natural language processing tasks.

## CHAPTER 6: SUMMARY AND CONCLUSIONS

This study captured more than 4300 annotations describing language used in the EMR to document patient reported symptoms and symptom changes. Of these annotations, almost 1200 were unique ways of describing the concept of symptoms for PTSD. We also captured more than 1700 annotations describing language used to document psychotherapeutic treatment modalities and treatment changes within the Veterans Health Administration (VHA) for PTSD. Of these annotations, more than 800 were unique ways of describing these treatment concepts. The level of concept extraction and the amount of unique data captured during this annotation task will allow researchers using NLP tools to produce more efficient and effective results in this clinical domain. Not to mention the contribution it will have to the creation of a PTSD ontology that can further enhance particular NLP tasks.

Specifically, NLP efforts are in process to use named entity recognition to identify the annotated concepts from this task. The annotated text and its corresponding concept identifier will be used to automate identification of the extracted dataset concepts on a much larger corpus of approximately 100,000 additional notes. Successful recognition of the named entities (concepts) allow for potential inference on a grander scale. Other NLP teams are creating tools to automate the categorization of our annotated psychotherapeutic and pharmacologic treatments and techniques. The ability to successfully generate categorization of concept-derived content in the clinical note not only reduces dimensionality and eliminates noise, but it is potentially useful for enhanced summarization of the course of clinical care for PTSD. One major and most promising effort will help address the nature of PTSD symptom expression

over time with respect to treatment introduction into patient care. This effort will create a tool to evaluate temporality language based on the annotated data from this study. The ability to successfully extract high level concepts and their associated relationship relative to time allows us to attempt to represent this information as a timeline of patient care and an expression of the disorder.

Ultimately, this work is intended to support the creation of clinical decision support systems for PTSD. Clinical decision support (CDS) requires a computer knowledge base of the language of the desired clinical domain to generate patient-specific assessments or recommendations to be presented to a provider.<sup>30</sup> Communication between the EMR and the CDS system is necessary for effective clinical decision making support systems. NLP essentially extracts the information needed from the EMR to actuate decision rules for the CDS system. Imagine how difficult and inefficient such a task would be in aiding the decision of providers working with patients suffering from PTSD without the language captured in our annotated dataset.

Finally, the manually annotated corpus of VA EMR notes will serve as a gold standard dataset and will be shared within the VHA within the secure VINCI environment. Thus, the dataset will be available for use to further research in the area of PTSD indefinitely.

## CHAPTER 7: REFERENCES

1. Seal KH, Metzler TJ, Gima KS, Bertenthal D, Maguen S, Marmar CR. Trends and risk factors for mental health diagnoses among Iraq and Afghanistan veterans using Department of Veterans Affairs health care, 2002-2008. *American Journal of Public Health*. 2009;99 (9):1651-1658.
2. *Treatment of Posttraumatic Stress Disorder: An Assessment of the Evidence*. Washington, DC: The National Academies Press, 2008. 1. Print.
3. Seal KH, Bertenthal D, Miner CR, Sen S, Marmar C. 2007. Bringing the war back home: Mental health disorders among 103,788 US veterans returning from Iraq and Afghanistan seen at Department of Veterans Affairs facilities. *Archives of Internal Medicine*. 167(5):476-482.
4. Batres A, Zeiss A. 2007 (January 16). Presentation to the Institute of Medicine Committee on the Treatment of Posttraumatic Stress Disorder, Meeting One: Treatment of PTSD in VA facilities and programs. Washington, DC: IOM.
5. Wilson JP, Friedman M, Lindy J, eds. 2001. *Treating psychological trauma and PTSD*. New York: The Guilford Press.
6. Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB. 1995. Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry* 52(12):1048-1060.
7. Lacy LW. *OWL: representing information using the Web Ontology Language*. Trafford Publishing, 2005.
8. South BR, Shen S, Jones M, et al. Developing a manually annotated clinical document corpus to identify phenotypic information for inflammatory bowel disease. *BMC Bioinformatics*. 2009;10(Suppl 9):S12. doi:10.1186/1471-2105-10-S9-S12.
9. Liu K, Chapman W, Hwa R, Crowley RS. Heuristic Sample Selection to Minimize Reference Standard Training Set for a Part-Of-Speech Tagger. *J Am Med Inform Assoc*. 2007;14(5):641–650. doi:10.1197/jamia.M2392.
10. Byrd RJ, Steinhubl SR, Sun J, Ebadollahi S, Stewart WF. Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records. *Int J Med Inf*. 2013. doi:10.1016/j.ijmedinf.2012.12.005.
11. Guo L, Zhang W, Chang S, Zhang L, Ott J, Wang J. MK4MDD: A Multi-Level Knowledge Base and Analysis Platform for Major Depressive Disorder. *PLoS One*. 2012;7(10). doi:10.1371/journal.pone.0046335.
12. Roberts A, Gaizauskas R, Hepple M, et al. Building a semantically annotated corpus of clinical texts. *J Biomed Inform*. 2009;42(5):950–966. doi:10.1016/j.jbi.2008.12.013.
13. South BR, Shen S, Barrus R, DuVall SL, Uzuner O, Weir C. Qualitative Analysis of Workflow Modifications Used to Generate the Reference Standard for the 2010 i2b2/VA Challenge. *AMIA Annu Symp Proc*. 2011;2011:1243–1251.
14. Waghlikar KB, Torii M, Jonnalagadda SR, Liu H. Pooling annotated corpora for clinical concept extraction. *J Biomed Semant*. 2013;4(1):3. doi:10.1186/2041-1480-4-3.
15. Mayer J, Shen S, South BR, et al. Inductive creation of an annotation schema and a reference standard for de-identification of VA electronic clinical notes. *AMIA Annu*

- Symp Proc AMLA Symp AMLA Symp.* 2009;2009:416–420.
16. Deleger L, Li Q, Lingren T, et al. Building Gold Standard Corpora for Medical Natural Language Processing Tasks. *AMLA Annu Symp Proc.* 2012;2012:144–153.
  17. Kipper-Schuler K, Kaggal V, Masanz J, Ogren P, Savova G. System evaluation on a named entity corpus from clinical notes. In: ; 2008. Available at: [http://www.lrec-conf.org/proceedings/lrec2008/pdf/764\\_paper.pdf](http://www.lrec-conf.org/proceedings/lrec2008/pdf/764_paper.pdf).
  18. Xia F, Yetisgen-Yildiz M. Clinical Corpus Annotation: Challenges and Strategies. In: ; 2012. Available at: [http://faculty.washington.edu/melihay/publications/LREC\\_BioTxtM\\_2012.pdf](http://faculty.washington.edu/melihay/publications/LREC_BioTxtM_2012.pdf).
  19. Chapman WW, Dowling JN. Inductive creation of an annotation schema for manually indexing clinical conditions from emergency department reports. *J Biomed Inform.* 2006;39(2):196–208. doi:10.1016/j.jbi.2005.06.004.
  20. Imler TD, Morea J, Kahi C, Imperiale TF. Natural Language Processing Accurately Categorizes Findings From Colonoscopy and Pathology Reports. *Clin Gastroenterol Hepatol.* 2013;11(6):689–694. doi:10.1016/j.cgh.2012.11.035.
  21. Savova GK, Chapman WW, Zheng J, Crowley RS. Anaphoric relations in the clinical narrative: corpus creation. *J Am Med Inform Assoc.* 2011;18(4):459–465. doi:10.1136/amiajnl-2011-000108.
  22. Rubin D, Wang D, Chambers DA, Chambers JG, South BR, Goldstein MK. Natural language processing for lines and devices in portable chest x-rays. *AMLA Annu Symp Proc AMLA Symp.* 2010;2010:692–696.
  23. Hazlehurst B, Sittig DF, Stevens VJ, et al. Natural Language Processing in the Electronic Medical Record: Assessing Clinician Adherence to Tobacco Treatment Guidelines. *Am J Prev Med.* 2005;29(5):434–439. doi:10.1016/j.amepre.2005.08.007.
  24. Meystre SM, Thibault J, Shen S, Hurdle JF, South BR. Texttractor: a hybrid system for medications and reason for their prescription extraction from clinical text documents. *J Am Med Inform Assoc.* 2010;17(5):559–562. doi:10.1136/jamia.2010.004028.
  25. Uzuner O, Solti I, Xia F, Cadag E. Community annotation experiment for ground truth generation for the i2b2 medication challenge. *J Am Med Informatics Assoc JAMIA.* 2010;17(5):519–523. doi:10.1136/jamia.2010.004200.
  26. Doan S, Kawazoe A, Conway M, Collier N. Towards role-based filtering of disease outbreak reports. *J Biomed Informatics.* 2009;42(5):773–780. doi:10.1016/j.jbi.2008.12.009.
  27. Pak A, Bernhard D, Paroubek P, Grouin C. A Combined Approach to Emotion Detection in Suicide Notes. *Biomed Informatics Insights.* 2012;5(Suppl. 1):105–114. doi:10.4137/BII.S8969.
  28. Pestian J, Pestian J, Pawel Matykiewicz, Brett South, Ozlem Uzuner, John Hurdle. Sentiment Analysis of Suicide Notes: A Shared Task. *Biomed Informatics Insights.* 2012;3. doi:10.4137/BII.S9042.
  29. Hripcsak G, Rothschild AS. Agreement, the F-Measure, and Reliability in Information Retrieval. *J Am Med Inform Assoc.* 2005;12:296-298.
  30. Demner-Fushman D, Chapman W, McDonald C. What Can Natural Language Processing Do For Clinical Decision Support? *J Biomed Informatics.* 2009;42:760-772. Doi:10.1016/j.jbi.2009.08.007.

## CHAPTER 8: TABLES

Item	Batch 1	Batch 2
Annotations (total)	2012	<b>2941</b>
Notes	585	<b>326</b>
Patients	572	<b>44</b>
VA sites	119	<b>6</b>

Table 1 Overall Difference between Batch 1 and Batch 2 Annotated Clinical Notes  
(Note: Batch 2 is more efficient)

Class	Batch 1 (total annotations)	Batch 2 (total annotations)
Symptom	782	<b>1528</b>
Symptom Change	47	<b>89</b>
Treatment	<b>447</b>	<b>501</b>
Treatment Change	10	1
Negation	54	<b>249</b>

Table 2 Batch Differences Based On Total Annotations for Specific Classes of Concepts Annotated in Clinical Notes

(Note: The lack of a significant difference in numbers for the treatment class due to only one instance of treatment being annotated per document)

Batch Number	Set Number	IAA Symptom	IAA Treatment
Batch 1	1	37.7 %	67.1%
	2	76.8%	67.0%
	3	42.9%	N/A
	4	59.4%	N/A
Batch 2	1	80.9%	81.7%
	2	81.3%	77.0%
	3	82.3%	65.9%
	4	81.2%	76.1%

Table 3 Inter-Annotator Agreement (IAA) for Batch 1 and Batch 2.  
 (Note: N/A for Set 3 and 4 of Batch 1 due to initial separation of notes into larger groups for only two sets.)

Batch	IAA (Average)
Batch 1	58.5%
Batch 2	78.3%
Class	IAA (Average)
Symptom	67.8%
Treatment	72.5%

Table 4 Average Inter-Annotator Agreement (IAA) for Batch and Class



<b>Class</b>	<b>Total annotations</b>	<b>Unique annotations</b>
<b>Symptom</b>	4051	<b>1111</b>
<b>Symptom change</b>	260	66
<b>Negation</b>	652	47
<b>Experiencer</b>	543	11

Table 5 Number of Total and Unique Annotations for Symptomatology Elements

<b>Class</b>	<b>Total annotations</b>	<b>Unique annotations</b>
<b>Treatment</b>	1706	<b>818</b>
<b>Treatment change</b>	12	12
<b>Treatment Non-compliance</b>	105	103
<b>Dosage</b>	154	42
<b>Duration</b>	459	96
<b>Frequency</b>	290	130
<b>Provider</b>	77	15
<b>Audience</b>	368	50

Table 6 Number of Total and Unique Annotations for Treatment Elements

Symptom Concept	# of Total Annotations	Treatment Concept	# of Total Annotations
Suicide	325	Therapy	555
Anxiety	289	Psychotherapy	291
Homicide	260	Cognitive Therapies (CPT,CBT)	241
Anger	197	<b>Identifying stuck points</b>	171
Depression	191	<b>Trauma account</b>	125
Avoidance	126	<b>Worksheet (homework)</b>	100

Table 7 Numbers of Most Frequently Annotated Concepts for Symptoms and Treatments

Treatment Category	Number of Unique Annotations
Psychological	452
Psycho-Education	160
Pharmacologic	75
Case Management	64
Psychosocial	36
Other	26

Table 8 Number of Unique Annotations for Each Treatment Category  
(Note: Data represented in Figure 6)

Symptom	Treatment
Difficulty concentrating	Supportive psychotherapy
Ineffective thinking	Socratic questioning
Anger	Cognitive mapping
Anxious	Develop rapport
Anxiety	Biofeedback
Intrusive memories/thoughts	CPT worksheets

Table 9 Example of Relationships between Documented Symptom and Treatment Prescribed

Number of Times Documented	Symptom Concept Targeted by Treatment Categories
24	Mood Instability
19	Nightmares
4	Suicidal Ideation
3	Panic
2	Crying

Table 10 Relationship between Documented Symptom Concept and Treatment Category Prescribed: Pharmacologic and Psychological

Number of Times Documented	Symptom Concept Targeted by Treatment Categories
47	Cognitive Distortions
19	Self-blame/guilt
18	Social Dysfunction
5	Triggers
4	Shame
3	Substance Abuse
2	Trauma Symptoms

Table 11 Relationship between Documented Symptom Concept and Treatment Category Prescribed: Psycho-Education and Psychological

Number of Times Documented	Symptom Concept Targeted by Treatment Categories
244	PTSD Symptoms
77	Anxiety
64	Stress
61	Depression
45	Insomnia
35	Anger
15	Agitation/irritability
13	Flashbacks
9	Avoidance
9	Hyper-arousal
9	Intrusive Memories/thoughts

Table 12 Relationship between Documented Symptom Concept and Treatment Category Prescribed: Pharmacologic, Psycho-Education and Psychological

## CHAPTER 9: FIGURES

The image displays a sample annotated document. On the left is a vertical list of classes with colored squares next to them: Begin\_Annotation (3) in green, Duration (3) in red, End\_Annotation (3) in red, Frequency (4) in blue, NarrativeContainer in tan, Negation (2) in cyan, PTSD\_Experiencer (4) in yellow, PTSD\_Symptom (26) in red, PTSD\_SymptomChange in purple, PTSD\_TreatmentAudience (2) in light green, PTSD\_TreatmentChange in cyan, PTSD\_TreatmentDosage in yellow, PTSD\_TreatmentNonCompliance in yellow, PTSD\_TreatmentProvider in light blue, PTSD\_TreatmentType (21) in blue, and TreatmentNote (1) in grey. On the right is a text document with several paragraphs. The text is annotated with colored highlights: '50 min.' is red, 'Individual' is yellow, 'PTSD; MDD' is red, 'Cognitive Processing Therapy (CPT)' is blue, 'PTSD' is red, 'patient' is yellow, 'challenging Beliefs' is blue, 'living/receiving a' is blue, 'commitment' is blue, 'each day' is blue, 'nightmares' is red, 'dreams' is red, 'could have done something to prevent him from dying' is red, 'problematic thinking and challenging questions' is blue, 'cognitive restructuring' is blue, 'screen-related stuck points' is red, 'stuck points related to intimacy' is red, 'stuck points' is red, and 'Impact Statements' is blue. The text also includes 'PROGRESS NOTE:', 'Content: This was the eleventh session of', 'for PTSD', 'completion of the', 'as well as', 'commitment', 'each day', 'Vet stated he thought he was getting past this, but then he had these', 'and he doesn't feel like he's getting over it. The Sgt. who died keeps appearing in his', 'Upon further exploration, vet still believes he', 'Therapist helped him to challenge this and', 'Examples from the worksheets were reviewed to offer further', 'and to fine-tune completion of the worksheets.', 'Stuck points', 'Stuck points were specifically targeted', 'Stuck points related to intimacy', 'stuck points', 'and he agreed to read materials related to this theme. The patient also agreed to complete a Challenging Beliefs worksheet about', 'each day, and to write another', 'describing his current thoughts and beliefs about himself, others, and the world related to his traumatic experiences.'

Figure 1 Sample Annotated Document

(Note: Schema to the left shows respective color of annotated text and corresponding class. Annotated text in red corresponds to PTSD\_Symptom class-Pass 1. Annotated text in blue corresponds to PTSD\_TreatmentType class-Pass 2.)

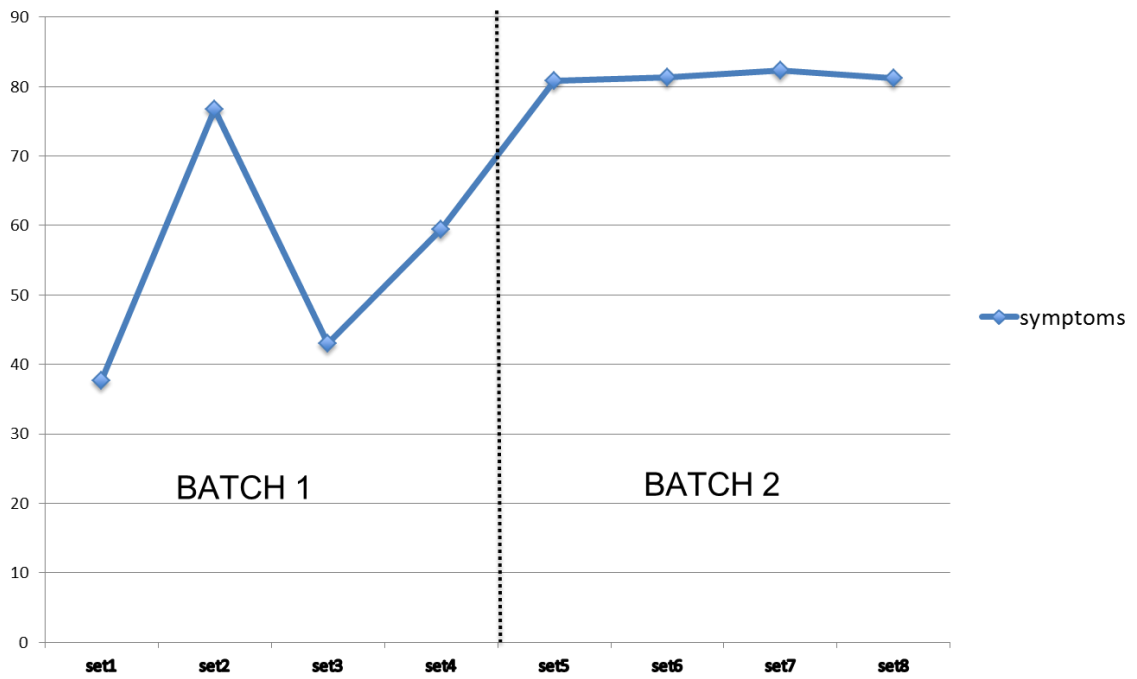


Figure 2 Evolution of Agreement for Symptom Pass

(Note: X axis=Annotated Document; Y axis=% Agreement; Blue line graph corresponds to symptom pass)



Figure 3 Evolution of Agreement for Treatment Pass

(Note: X axis=Annotated Document; Y axis=% Agreement; Red line graph corresponds to treatment pass)

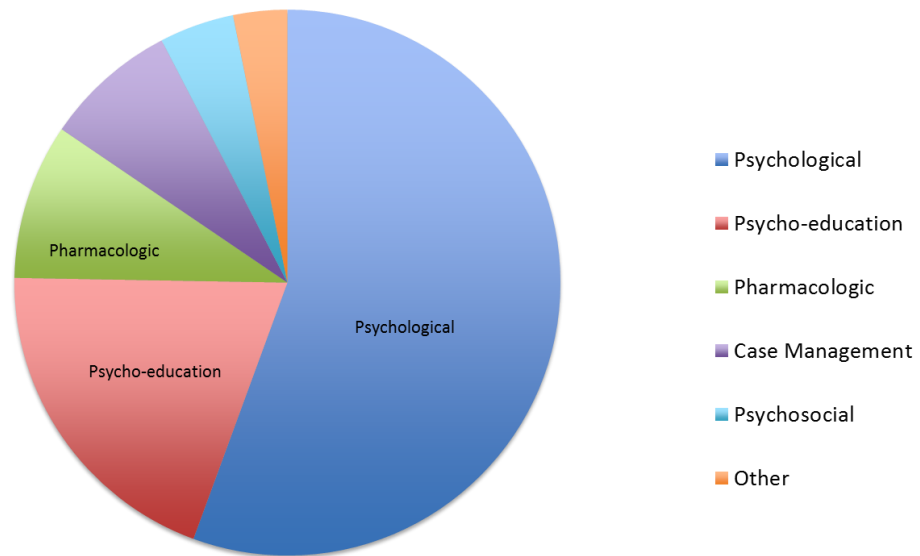


Figure 4 Numbers of Unique Annotations for each Treatment Category

(Note: Color and Treatment Category correspond to color in pie chart)

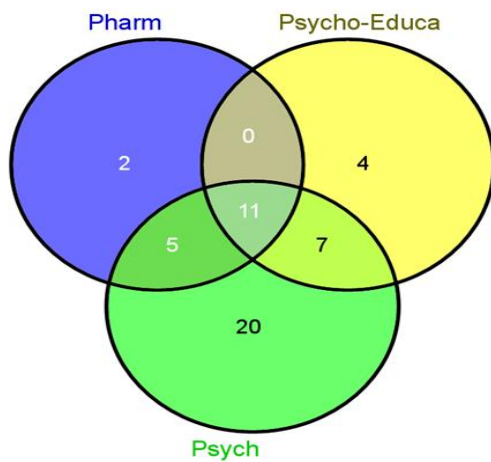


Figure 5 Venn Diagram for Relationship Between Documented Symptom Concept and Treatment Category Prescribed: Pharmacologic, Psycho-Education, and Psychological



## CHAPTER 10: APPENDIX

### CHIR PTSD Annotation Guidelines

Post-Traumatic Stress Disorder (PTSD) is a mental health disorder affecting a great deal of United States veterans. The psychological treatment received by veterans is not adequately captured within structured data in the VHA electronic health record. In most cases, it is only accessible through the narrative portion of a clinical note. As a part of the CHIR initiative to develop methods and tools for effective use of unstructured data such as narrative text and apply these tools to address specific research questions, we need to first examine the information existing in the clinical notes, which captures encounters with veterans suffering from PTSD.

The criteria in this guideline are based on DSMIV (found in CHIR PTSD Annotation Guidelines Appendix) and NAMI for the operational definitions of PTSD symptoms and treatment.

For annotation tasks we will use Knowtator on [sourceforge:http://sourceforge.net/projects/knowtator/files/Knowtator](http://sourceforge.net/projects/knowtator/files/Knowtator). The Knowtator tool is written for Protégé 3.4.8. This version of Knowtator includes specific modifications to make the annotation task faster and more efficient. See CHIR PTSD Annotation Guidelines Appendix for the schema for concept extraction annotation. **NOTE: Concepts for annotation listed in this guideline are color coded to match the schema used during annotation.**

#### I. General Information

##### 1. Annotation Scope

This project is composed of several distinct guidelines, one for each concept. The concepts of interest have been clearly defined in each annotation task: symptoms pass and treatment pass. For all annotations in all classes, the scope is limited to information directly representing the concepts of each guidelines set. The definitions have been made as explicit as possible, but some clinical judgment and context must be used to determine if the text is one of the desired concepts. Many synonyms and modifications exist for each of the concepts, and reviewers should annotate these variations if they meet the strict definition.

## 2. Assertions

The default assertion for these tasks is positive and related to the patient. Unless otherwise stated, only annotate mentions that are present/confirmed AND related to the patient. For those cases where items are negated, annotation must correspond to the negation slot value as requested. If associated with someone else, annotation must correspond to the experiencer slot value as requested. Hypothetical (not present but might happen in the future) mentions are excluded. When an annotation is made, its default status is positively asserted. Therefore most annotations will have the “assertion” slot left blank to indicate it was positively asserted. In rare cases, annotations may not be explicit enough to warrant certainty and require the mark of “possible” in the “assertion” slot. Positive (the default) and possible (selected in the slot) are the only possible assertions in this task. For the most part, annotation should only occur if the annotator is certain. Otherwise, do not annotate and request guidance from experts.

## 3. Annotation span

Spans must be one continuous string of text and may not be bridged when unmarkable text lies within. The span should be the minimal text span that captures the concept.

## II. Technical Information

### 1. Treatment Note (TreatmentNote)

For the treatment pass only, read through the note to determine if the specific note describes a treatment for PTSD. If it does have information about a treatment process, then assert ‘yes’ as a slot value drop down and move on to **Begin\_annotation**. If it does not, then assert ‘no’ as a slot value drop down and move on to **End\_annotation**. Please annotate something that is indicative of the note being a treatment note. Or, in the event that it is not a treatment note, annotate something that might summarize what kind of note it is.

### 2. Beginning annotation (**Begin\_annotation**)

For both symptom and treatment passes, annotate the first word in the note after deciding it is a treatment note. Do not annotate for **Begin\_annotation** if you determine it is not a treatment note. After asserting **Begin\_annotation**, annotate for clinical information. If the first word for **Begin\_annotation** satisfies another

annotation category and can be found elsewhere in the note, do not annotate anything other than **Begin\_annotation**.

**Example:** *“Patient saw the therapist for...*

*Patient is elsewhere in the note, only annotate as **Begin\_annotation** and look for the next instance of “patient” to fulfill the **PTSD\_Experiencer** category. Follow this same rule for the category of **End\_annotation**.*

### 3. Ending annotation (**End\_annotation**)

For the symptom and treatment passes, use this annotation after annotating for clinical information OR after deciding it is not a treatment note. The last word should be annotated for category **End\_annotation**.

**Exception:** *If it is a name, take the last word before the name. If it is a date, take the last word before the date. NEVER ANNOTATE (Personal Health Information) PHI. Make sure you are scrolling to the end of the note before you commit to your **End\_Annotation**.*

### 4. Relationships between classes of information

This section specifies how to create relationships between classes of information.

Relationships are indicated using the relationship slot attribute. The proper way to assign a relationship requires the text for each annotation in the relationship to be annotated first and then given a class designation before the relationship is to be made using the slot value attribute.

**Example:** *“Therapist supplied patient with CPT worksheets to help work through intrusive memories”*

First annotate for:

*CPT worksheets = PTSD\_TreatmentType*

Then annotate for:

*Intrusive memories = PTSD\_Symptom*

Create a relationship by using the slot value of **PTSD\_Symptom** and click the



*(add an instance button)* and select the required treatment of *CPT worksheets*. You will see the designated text appear in the slot value window if done correctly.

## 5. Other

- When annotating the term “*veteran*” for the category **PTSD\_Experiencer**, do not use that same instance of the word “*veteran*” for the category **PTSD\_TreatmentAudience**. Please find another instance of the term “*veteran*” to annotate for the other category. (The same word at a different location in the note is okay, the same word at the same location is not.) The adjudication system makes you choose which one to accept. It is not possible to have both. When this happens we lose data or we have to re-annotate for that category.
- Be aware that you are listed as the annotator on each annotation, and that your name is listed at the top as the filtered view.
- Be aware of repetitive text that is most likely a template inserted into the clinical document. If this is specific to the individual in any way, use judgment as to any symptoms or treatment that would fall under our concepts for annotation. In some cases, these templates will not be specific to the patient and will not require any annotations.

## III. Clinical Information

### Pass 1: Symptom Concept Annotation

#### 1. Symptom (**PTSD\_Symptom**)

Symptoms are organized according to diagnostic cluster (DSM-IV diagnosis criteria). This is done in post processing by the adjudicators.

We are only concerned with any mention of symptoms related to PTSD. Either asserted or negated. Annotate every instance of a symptom even if documented several times in the same note. In an effort to capture only those symptoms related to PTSD, do not annotate any diagnoses of depression or MDD as a symptom of PTSD. Only annotate “*depressive symptoms*” or “*depression*” if you are certain it is not related to another mental health disorder. This category may include terms such as: *anxiety, fear, guilt, and crying*

## 2. Negation (Negation)

Providers document negation of PTSD-related symptoms. This is a relationship and should be created as such following **II (4) Relationship between classes of information**. Annotate only those mental health symptoms that are explicitly denied in the text. Make sure that you are **not** capturing an annotation for the category **PTSD\_TreatmentNonCompliance**.

## 3. Symptom Change (PTSD\_SymptomChange)

This is an indication of any PTSD symptom changing for the better or for the worse. This is a relationship and should be created as such following **II (4) Relationship between classes of information**.

*Example: "Increase in anxiety" or "much less depressed"*

## 4. Experiencer (PTSD\_Experiencer)

This information includes the person being referenced for said symptom. This should be the first mention of the "patient" or "veteran" within the note. If patient or veteran does not exist, annotate "she" or "he." This is a relationship and should be created as such following **II (4) Relationship between classes of information**. The experiencer may be the patient or someone else. Some notes might reference a symptom or a change in symptom where the person describing the symptom is not the patient. The experiencer may be the patient's spouse or another patient in a group setting, but will be discussed in the note.

## 5. Symptom Frequency (Frequency)

This information includes the frequency of symptom occurrence. This is a relationship and should be created as such following **II (4) Relationship between classes of information**.

## **Pass 2: Treatment Concept Annotation**

### **1. PTSD Treatment Type (PTSD\_TreatmentType)**

Annotate concepts or phrases that meet the strict definition of one of the six treatment types below. Only annotate mentions that occurred during this encounter. Annotate this instance only once, even if documented several times during this encounter. Include any abbreviations or acronyms that represent the treatment types or strategies within these types below ONLY if there is certainty that they represent one of those concepts AND meet the strict definition. Any questionable abbreviations or acronyms should be excluded. When in doubt about any concept, phrase, abbreviation, or acronym please exclude it or bring it to the attention of the expert. When annotating for a type of therapy, take “individual therapy” in its entirety. If there are two treatments separated by the word “and” annotate as one treatment if they are clearly similar, but annotate separately if they are different. If a treatment type is mentioned more than once in the narrative, only annotate the first mention. After annotating text for this class you will have the option to create relationships, as usual, using slot values. The slot value for treatment type will be at the bottom of the slot value menu and will provide the available options from which to categorize the treatment. The definition for these options is listed below:

**Case Management**-Any collaborative process of assessment, planning, facilitation and advocacy for options and services to meet an individual’s holistic needs through communication and available resources that promote quality, cost-effective outcomes.

***Example:** treatment planning and resources*

**Psycho-education**–Any therapy that involves teaching patients about their illness, how to treat it, and how to recognize signs of relapse so that they can get necessary treatment before their illness worsens or occurs again. This includes teaching coping strategies and problem-solving skills. This excludes any educational CBT techniques that should fall under the category of psychological.

**Pharmacotherapy**–Any medication prescribed or any over the counter medication currently taken or offered for consideration in the future within the confines of the therapeutic relationship for the treatment of symptoms should be annotated under pharmacotherapy category related to PTSD (depression, anxiety, sleep problems, etc.)

Do not annotate titles or headings on templates (e.g. Active Medications).

Annotate any mention of medications related specifically to the management of PTSD in the narrative text. If you are unsure which medications might be used for PTSD, it is best to Google the medication to see what it is most typically used for, in the event that it is not documented in the note. If the same medication is also mentioned in the medication list, **do not** annotate. However, if it is only mentioned in the medication list, **do** annotate.

Make sure to annotate the entire medication in the med list or in the narrative.

**Psychosocial** – Any combination of psychotherapy and social and educational/vocational training

*Example: Recreational and social interactions technique*

**Psychological**- involves regular sessions with mental health providers to help patients understand why they are acting and thinking in ways that are troubling or dangerous to themselves or others so they have more control over their behaviors and can correct them. This includes therapies such as: CBT, CPT, DBT, trauma focused therapy, etc. This can also include general therapy techniques such as building a rapport, active listening. However, this will not include typical generic communication language such as praise or encouragement that is not tied to a specific therapeutic intervention.

**Other**- involves any text related to treatment for PTSD that does not fulfill the other five categories.

*Example: Chapel services, admission and restraints.*

## 2. Targeted Symptom (PTSD\_Symptom)

Symptoms are organized according to the use of the symptom in the note, and by diagnostic cluster (DSM-IV diagnosis criteria). Patient experiences a PTSD symptom related to the treatment annotated. This is a relationship and should be created as such following **II (4) Relationship between classes of information**. Do not annotate any symptom unless it is related to a treatment being annotated in the text. In the instance of a changing symptom related to a specific treatment, this annotation should be made in its respective change in symptom slot value. There should only be symptoms as an instance of treatment. No stand alone categorical symptoms should be annotated in this pass. **Note: a PTSD\_SymptomChange not connected to a specific treatment should only be annotated in Task 1.**

### 3. Treatment Change (PTSD\_TreatmentChange)

This is a change in treatment discussed within the note that is currently taking place. This is **not** a hypothetical reference to a change in treatment. Annotate text indicating a change even if the dose isn't specified, (e.g. increase/decrease medication). Annotate the change and the medication. If note is unclear as to whether a patient is on a medication or not or it has not been specifically discontinued (d/c) do not annotate the medication.

### 4. Audience (PTSD\_TreatmentAudience)

Group- Three or more individuals who are not related

Family- Two or more individuals who are related

Couples- Two or more individuals who are in a relationship (dating/married/separated/divorced)

Individual- One individual

This class is limited only to positively asserted settings for PTSD treatment management. This is a relationship and should be created as such following **II (4) Relationship between classes of information**. If the type of treatment is part of the same phrase where the setting is written you must annotate it under audience.

*Examples: individual, group, and five group members.*

### 5. Experiencer (PTSD\_Experiencer)

This information includes the person being referenced for said treatment or response to treatment. This should be the first mention of the “*patient*” or “*veteran*” within the note. If patient or veteran does not exist, annotate “*she*” or “*he*”. This is a relationship and should be created as such following **II (4) Relationship between classes of information**.

### 6. Treatment Duration (Duration)

Make sure that the allotted time is for the entire treatment.

*Example: 60 min of Individual therapy.* The 60 min can be annotated as a categorical value and a slot value for “individual therapy”.



**Example:** 30 min of supportive therapy and CBT skills training. The 30 min cannot be counted as a categorical or slot value for either treatment as we are not sure how long each took to administer.

## 7. Treatment Providers (PTSD\_TreatmentProvider)

Clinician- MD, DO, NP, PA

Practitioner- PhD, PsyD, EdD

Case Manager- MSW, LCSW

Therapist- LCT, LMFT, LMHP, LPC

Staff- RN

This information is usually found in the signature of the note, if it is not referenced within the text of the note. Be careful to only annotate a provider instance that is explicitly mentioned for the current treatment encounter.

## 8. Treatment Dosage (PTSD\_TreatmentDosage)

This information includes specific dosage of a medication, or specific time spent using a treatment technique that does not fall under duration of treatment encounter.

## 9. Treatment Frequency (Frequency)

This information includes the frequency of medication or psychotherapy sessions.

## 10. Noncompliance (PTSD\_TreatmentNoncompliance)

This information includes a comment by the provider or patient that a certain treatment has not been used or completed in compliance with the patient's treatment plan.

**Example:** patient is not taking their prescribed medications and no show.

#### IV. Appendix



#### DSM-IV-TR Post-Traumatic Stress Disorder Diagnostic Criteria

A. The person has been exposed to a traumatic event in which both of the following were present:

1. the person experienced, witnessed, or was confronted with an event or events that involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others
2. the person's response involved intense fear, helplessness, or horror. Note: In children, this may be expressed instead by disorganized or agitated behavior

**B.** The traumatic event is persistently re-experienced in one (or more) of the following ways:

1. recurrent and intrusive distressing recollections of the event, including images, thoughts, or perceptions. Note: In young children, repetitive play may occur in which themes or aspects of the trauma are expressed.
2. recurrent distressing dreams of the event. Note: In children, there may be frightening dreams without recognizable content.
3. acting or feeling as if the traumatic event were recurring (includes a sense of reliving the experience, illusions, hallucinations, and dissociative flashback episodes, including those that occur on awakening or when intoxicated). Note: In young children, trauma-specific reenactment may occur.
4. intense psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event
5. physiological reactivity on exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event

**C.** Persistent avoidance of stimuli associated with the trauma and numbing of general responsiveness (not present before the trauma), as indicated by three (or more) of the following:

1. efforts to avoid thoughts, feelings, or conversations associated with the trauma
2. efforts to avoid activities, places, or people that arouse recollections of the trauma
3. inability to recall an important aspect of the trauma
4. markedly diminished interest or participation in significant activities
5. feeling of detachment or estrangement from others
6. restricted range of affect (e.g., unable to have loving feelings)
7. sense of a foreshortened future (e.g., does not expect to have a career, marriage, children, or a normal life span)

**D.** Persistent symptoms of increased arousal (not present before the trauma), as indicated by two (or more) of the following:

1. difficulty falling or staying asleep
2. irritability or outbursts of anger

3. difficulty concentrating
4. hypervigilance
5. exaggerated startle response

**E.** Duration of the disturbance (symptoms in Criteria B, C, and D) is more than 1 month.

**F.** The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

**Specify if:**

Acute: if duration of symptoms is less than 3 months

Chronic: if duration of symptoms is 3 months or more

**Specify if:**

With Delayed Onset: if onset of symptoms is at least 6 months after the stressor.