How Hospitalized COVID-19 Patients Vulnerable to Trauma Narrate Their Experience: A Computational Linguistic Analysis of COVID-19 Patients' Experiences during and after Hospitalization

Marieke van der Bie, 2632292
Department of Psychology, University of Twente

202200087, Master thesis Positive Clinical Psychology & Technology

First supervisor: Dr. Anneke Sools

Second supervisor: Prof. Dr. Harald Vonkeman

July 8, 2024

Abstract

The COVID-19 pandemic, which started in late 2019, had caused many infections and deaths worldwide by November 2023. Not only the physical effects, but its psychological impact on hospitalized patients with the development of Posttraumatic Stress Disorder (PTSD) has gained significant attention. This study explores the language use of COVID-19 patients, to understand their experience of trauma. The study used a mixed-method approach combining quantitative text mining with qualitative interpretation by analysing the narratives from 36 patients scoring low and 36 patients scoring high on PTSD symptomatology. These patients were admitted to the ZGT and MST hospitals in The Netherlands due to their COVID-19 infection. The patients provided two stories: their admission story and their recovery story, along with completing the PCL-5 questionnaire on PTSD symptomatology. The analysis consisted of a comparison of language use (topics and sentiment) in the two stories of high and low PTSD symptomatology, complemented by qualitative interpretation. The results have revealed distinct differences between the subgroups. Patients with low PTSD symptomatology show a positive and accepting perspective toward their hospital admission, focusing strongly on procedural knowledge and expressions of optimism about their recovery. Conversely, patients with high PTSD symptomatology show emotional struggles and isolation. Their narratives are embedded with negative sentiments that reflect pessimism and distress. The findings can serve as a foundation for future research, underscoring the importance of clear information provision at the start of admission, social support, and considering comorbidities.

Keywords: COVID-19, posttraumatic stress disorder, trauma, hospitalization, computational text mining

Table of Contents

1. COVID-19 Patients' Experiences during and after Hospitalization

The *COVID-19* pandemic, which started in late 2019, rapidly developed into a global crisis by early 2020. By November 2023, the SARS-CoV-2 virus, also known as the COVID-19 virus, had caused over 772 million infections and 6.9 million deaths worldwide. Of those, 8.6 million illnesses and 23.000 deaths occurred in The Netherlands (Dong, Du & Gardner, 2020). The virus was first considered a significant international public health concern, which was later corrected by the World Health Organisation (WHO), as the pandemic developed (WHO Coronavirus (COVID-19) Dashboard, n.d.). The virus is primarily transmitted through airborne particles which led to worldwide isolation measures. The symptoms after an infection can range from flu-like symptoms to severe respiratory distress. Although hospitalization rates have currently declined, the impact on our health has become visible. Patients experience neurological symptoms, like loss of taste and smell, and headaches. Others still experience shortness of breath or fatigue (Helms et al. 2020; Mao et al., 2020).

Apart from the physical symptoms, the psychological consequences of hospitalization, isolation during treatment, and uncertainty about the illness contribute to feelings such as fear, anger, and loneliness (Xiang et al., 2020). Studies revealed that survivors of COVID-19, hospitalized or not, suffer from psychological symptoms, including depression, anxiety, and insomnia (Liu et al., 2021; Vanderlind et al., 2020; Zhang et al., 2020). Moreover, Penninx and colleagues (2022) found that symptoms like depression and fatigue can last up to several months after the infection.

The recovery journey after hospital admission can be an intense experience, resulting in both physical and psychological changes. Research suggests that patients struggling with life-threatening illnesses, including COVID-19, may experience symptoms associated with *Post-Traumatic Stress Disorder*[1] (PTSD), even after being fully physically recovered (Xiao et al., 2020). Concerning *Intensive Care Units* (ICU), Tedstone and Terrier (2003) found prevalence rates of PTSD symptoms ranging from 14-59%. Recent studies by Bo et al. (2021) and Chamberlain et al. (2021) also confirm the occurrence of PTSD symptoms in clinically stable COVID-19 patients, with intrusive images like visual memories being the main symptom. In comparison to the SARS-CoV-1 outbreak in 2003, the COVID-19 pandemic demonstrates a higher likelihood of PTSD symptoms, introducing concerns for overall

[1] The DSM-V describes PTSD as involving symptoms like avoidance, sleep disruptions, distressing dreams, irritability, and headaches, ultimately leading to disruptions in social functioning and perceived quality of life (American Psychiatric Association, 2013).

healthcare (Vindegaard & Benros, 2020). These risks include chronic mental health issues, like major depressive disorder and substance abuse (Kessler et al., 1995).

Beyond the specific illness, the time after the hospitalization itself is a key factor in developing PTSD symptoms. Research on hospitalization and traumatic medical experiences in general highlights the prevalence of post-hospitalization trauma (Shih et al., 2010; Zatzick et al., 2007). An ICU hospitalization can be traumatic for several reasons: isolation, loss of control, and invasive medical procedures. Chahraoui and colleagues (2015) found that hospitalized patients experience feelings of powerlessness due to isolation, as well as having a near-death experience that causes anxiety and (repetitive) nightmares. PTSD symptoms show significant prevalence among ICU survivors (Davydow et al., 2008; Ratzer et al., 2014; Righy et al., 2015). Regarding the nature of PTSD symptoms, early identification is important as trauma can lead to physical complications, including chronic pain and hypertension (Zatzick et al., 1997). Undiagnosed PTSD may even increase the risk of suicide (Stanley et al., 2017), making timely interventions crucial for preventing prolonged personal suffering among patients hospitalized due to COVID-19.

Several measurement options exist to assess the presence (and severity) of PTSD, like structured interviews and self-reported questionnaires. A commonly used measure is the PTSD Checklist for DSM-5 (PCL-5), which assesses symptoms based on the criteria defined in the DSM-5. This checklist contains 20 items that match the DSM-5 criteria like intrusion, avoidance, and mood of the patient. A few other measurements are available for assessing PTSD symptomatology, like the Clinician-Administered PTSD Scale for DSM-5 (CAPS-5). This is a structured clinical interview, conducted by a qualified or trained clinician to evaluate the presence and severity of PTSD. While such measures are valuable in evaluating symptomatology, several factors can complicate the assessment process. PTSD manifests differently in different patients and is influenced by a variety of factors. For example, it may depend on the nature of the event, cultural background, or coping mechanisms. On top of that, comorbidities can exist, such as depression or anxiety, that overlap with PTSD symptoms.

Literature on PTSD and language use is sparse, offering limited insights into how these individuals express their traumatic experiences through linguistic characteristics. Still, Paquet et al. (2024) highlighted the implications of language use in Posttraumatic Nightmares (PTNM)[2] on psychological symptoms, addressing a need for further exploration into language

[2] Post-traumatic nightmares are recurrent frightening dreams that happen after a traumatic event (Sandman et al., 2017). These nightmares can differ in content, with some covering the trauma directly, while others may be in symbolic form (Nielsen, 2017).

use in PTNM narratives. On the other hand, Papini and colleagues (2005) investigated linguistic characteristics beyond trauma narratives. They found that individuals with PTSD showed linguistic patterns, including an increased use of singular pronouns (e.g., “I” and “me”), death-related words, and a decreased use of plural pronouns (e.g., “we” and “us”). This insight suggests a more self-focused, mortality-aware writing style that is less concerned with collective experiences. The severity of specific PTSD symptoms correlated with variations in linguistic markers, making language a strong predictor of PTSD.

Furthermore, PTSD and language use have often been studied in the context of memory retrieval and disorganization in indicating PTSD (Jelinek et al., 2010). Myers et al. (2022) focused on patients with mild Traumatic Brain Injury (mTBI) and PTSD, using discourse analysis to assess cognitive communication skills. The study showed differences in global coherence and word count, suggesting shortcomings in verbal fluency. However, other studies could not find support for this relationship between PTSD and disorganized language use (Gray & Lombardo, 2001; Peace et al., 2008). For a long time, literature has indicated that there is not enough evidence about the organization of trauma narratives in people with PTSD to offer any conclusion about whether PTSD narratives are well-organized or fragmented (O’Kearney & Perrott, 2006).

1.1 Using Computational Text Mining to Uncover Meaning within Self-Reported Patient Narratives

The patient narrative shows to be promising concerning exploring the mental health of patients post-COVID-19. Language, an important method to express thoughts and emotions, can help reveal mental health problems (Atapattu et al., 2022). Early linguistic development as a toddler has been linked to better adolescent mental health, highlighting the significance of language in social interaction and emotional well-being (Thornton et al., 2021). Large-scale resources for exploring (online) language use related to mental health conditions seem essential for gaining insights into mental health. These resources could provide datasets important for making use of this narrative to understand the experience of individuals with increased PTSD symptomatology (Cohan et al., 2018). The shortage of large, shareable datasets based on actual medical grounds is troublesome because of the need for high-quality, clinical-based datasets for mental health research (Harrigian et al., 2020). Additionally, the time of diagnosis provided signals for examining mental health through language, especially in self-reports (MacAvaney et al., 2018). Not only patients themselves but also their caregivers’ attitudes can be studied. For example, programs engaging with mental health service users to share their recovery narratives can reduce stigma and facilitate healing, while also promoting engagement in mental health care (Kaiser et al., 2020).

Narrative studies that focus on both the content and structure of narratives, can provide a unique perspective to explore the experiences of hospitalized patients. They can act as a powerful tool for uncovering the experience while also constructing meaning. Research by Bury (2001), highlights the importance of narratives in illness, emphasizing how individuals use storytelling to make sense of their health challenges. Additionally, it suggests that writing stories allows for a deeper exploration of how patients interpret and navigate their mental health conditions. Furthermore, Hydén (1997) studied the role of narratives in the context of illness and recovery, providing insights on how storytelling serves as a way for individuals to express emotions, and find coherence in their suffering. Therefore, this study aims to make a distinction between the hospitalization story, and the recovery story.

By employing *Computational Text Mining* techniques, a structured method to help study language use is provided. Text mining, as explained by Trusko et al. (2010), involves the alteration of unstructured text into structured data, allowing for the identification of relevant information. As defined by He, Veldkamp, and De Vries (2012), text mining seeks patterns in unstructured textual data, revealing insights that may not have been found by solely using traditional qualitative analyses. The potential of text mining is found in Howes and colleagues' (2014) study, where topics and sentiment in patient narratives were able to predict depressive symptom severity and (partly) patient progress.

Computational text mining technology offers multiple advantages. It can enhance the efficiency of processing structured texts through pre-processing and identify (hidden) patterns not accessible through traditional qualitative analysis (Gupta & Lehal, 2009; Ledeneva & Sidorov, 2010). On top of that, in psychology, computational text mining provides objective measurements rather than subjective human interpretation (Short et al., 2018). Specific text mining techniques like topic and sentiment analysis can indicate subtle linguistic cues indicative of trauma-related symptomatology (Sawalha et al., 2022). However, human interpretation will remain crucial in understanding the use of language in mental health-related narratives. Researchers in particular, guide the computer in tasks such as pre-processing and aiding contextual understanding, empathy, and recognising emotional cues that cannot be found through computational analysis alone.

1.2 The Current Study

While previous narrative research made progression in making sense of narratives of those with persistent symptoms (long COVID[3]), a gap in understanding the experiences of patients specifically hospitalized due to COVID-19 continues, for example, ICU admissions. Prior studies, exemplified by Ladds and colleagues (2020) and Rushforth and colleagues (2021), laid foundations by exploring illness experiences and the structure of narration. Their insights, however, did not necessarily focus on those hospitalized patients – an area of research that awaits thorough investigation.

This study aims to describe and understand the use of language of the patients hospitalized due to COVID-19, and indirectly their experience of trauma. By analysing the divergences and convergences in language use between the experience of hospitalization and recovery, the study aims to provide meaningful insights for interventions and precautionary measures in hospitals and other instances, contributing to a richer understanding of PTSD experiences in existing literature and hospitalized patients' needs. In this regard, this study will explore the following research questions: How do trauma-related narrative responses, in terms of sentiment and topics, relate to PTSD symptomatology in patients who have been hospitalized due to COVID-19? And what similarities and differences exist between language use during the hospitalization and recovery stage?

2. Methodology

2.1 Design

The current study used a mixed-method approach to text mining that involves both statistical techniques and qualitative interpretation of linguistic patterns in the narrative responses of patients with high and low PTSD symptomatology. The study consists of a comparative analysis, comparing two groups (low and high PTSD symptomatology) and the comparison between stages (hospitalization and recovery). The study is part of a larger longitudinal study conducted by the University of Twente in collaboration with MST hospital (Enschede), of which data was gathered from two hospitals in The Netherlands: the MST in Enschede and ZGT in Almelo. The MST hospital aims to enhance patients' well-being after discharge, as well as to be more prepared for future pandemics or health crises.

---

[3] Long-Covid, also known as post-acute sequelae of SARS-CoV-2 infection (PASC), refers to the condition where individuals continue to experience symptoms and signs associated with COVID-19 for more than three months after the initial infection (Seo et al., 2024).

2.2 Participants

The respondents were Dutch patients being hospitalized at MST and ZGT from November 2020 until August 2022. Up to one year after discharge, the patients completed a set of questionnaires. The inclusion criteria for this study included being above the age of 18 years old, hospitalized with PCR-proven[4] COVID-19, having completed the PCL-5 questionnaire about PTSD symptomatology, and two narrative questions in which patients are asked to tell their hospitalization and recovery story. This resulted in a total of 450 patients that filled in both questionnaires. However, several missing values were present. For example, missing demographic data, or patients that filled in “Nee” [No] or “x” for both narrative questions. This resulted in a total of 432 patients that were suitable to use for analysis.

Based on research, the researchers chose to use a specific cut-off score for the PCL-5 to determine PTSD symptomatology. Praag and colleagues (2020) conducted a validation study of the Dutch version of the PCL-5 in a civilian population after traumatic brain injury. Additionally, Bockhop and colleagues (2022) examined the measurement invariance of the PCL-5 in individuals, offering information on the factorial validity and cross-linguistic equivalence of the questionnaire. Tohme et al. (2024) highlighted the variability in cut-off scores, ranging from 28 to 37, underscoring the need to consider this range. A cut-off score of 28 is used to ensure the use of as much data as possible.

In this study, we aimed to investigate the distribution of scores on the PCL-5 among the sample population. Upon analysing the PCL-5 scores, a wide range of responses were observed, from minimal to severe symptomatology. To facilitate a clearer understanding of the data, and to categorise individuals based on their symptom severity, it is crucial to create subgroups. Given the distribution of scores and the complexity of PTSD as a disorder, it was decided to include patients that it have a score of at least 28 on the PCL-5 questionnaire to maximize contrast between subgroups, resulting in 36 patients. This group of patients has a mean age of 57,4 (*SD* = 9,9, range 33 to 77), 22 males (61,1%) and 14 (38,9%) females. In this group, the PCL-5 questionnaire had a mean of 36,5, which means a high likelihood of the occurrence of PTSD (*SD* = 8,8, range 28 to 64).

The low-scoring group also consisted of 36 patients. In the participant selection process, patients who scored 0 on the PCL-5 test were excluded from the analysis. This decision aimed to ensure the reliability of the collected data by minimising the potential response bias. The inclusion of patients who may not have provided truthful responses could

[4] Chi et al. (2023) “The gold standard of COVID-19 diagnosis is polymerase chain reaction (PCR)” (p. 2)

introduce bias into the data, influencing the reliability of the findings. The selected patients matched the high-scoring group in demographics and length of narratives. The 36 patients that scored highly on the PCL-5 questionnaire, had a mean age of 60,9 (*SD* = 12,8, range 35 to 79), 21 males (55,0%) and 15 (45,0%) females. In this group, the PCL-5 questionnaire had a mean of 2,8, which means that there is a likelihood of no occurrence of PTSD (*SD* = 1,5, range 1 to 5).

2.3 Measurements

At 6, 9, and 12 months after discharge, the patients filled out questionnaires about their clinical and psychosomatic symptoms that were sent to them by email. For this study, data was available at 6 and 12 months after discharge. These questionnaires consisted of several parts. One of those were the narratives in response to open-ended questions: “Kunt u het verhaal vertellen over uw opname in het ziekenhuis vanwege het Coronavirus?” [Could you tell the story about your hospital admission because of the coronavirus?], and “‘’Kunt u het verhaal vertellen over hoe uw leven na uw opname is verlopen?’’ [Could you tell the story about your life after the hospital admission?]. The patients could write what came to their minds without a minimum or maximum of words.

Next to the narratives, the patients filled out the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5) questionnaire. The PCL-5 indicates PTSD symptomatology (Boeschoten et al., 2020).

2.4 Data Analysis

2.4.1 Software

The first software used is RStudio (R Version 2023.09.1+494), an integrated development environment (IDE). R is an open-source software widely used for its capabilities in data analysis, statistical modelling, and visualisation, as well as for its flexibility and functionalities. It contains a range of packages, making it well-suited for a range of analytical needs, also used for this study (e.g., package dplyr, see Appendix A). The initial step involved importing and pre-processing raw data, as well as the analysis of demographic data. R facilitated such tasks as loading datasets, handling missing values, and recoding variables, ensuring that the data were appropriately prepared for the analyses that followed.

After pre-processing the data, text mining was executed using the Orange3 sofware. Orange3 is an open-source data visualisation and computational analysis tool that is focused on handling complex tasks, including language processing and text analysis. It was selected for its user-friendliness as well as its functionalities for text mining. Its capabilities in handling textual data, particularly topic modelling and sentiment analysis, made it a well-suited choice for this study.

2.4.2 Pre-processing

As the study aims to compare language use, indirectly trying to understand the experience of trauma, the first pre-processing task was to remove short answers that did not give direct insights into this experience. For example, one patient answered both narrative questions with: “Ja” [Yes], and another patient answered: “Reeds verteld” [Already told before]. However, some answers were more difficult to assess based on their meaning. According to Thelwall and Buckley (2013), it is important to consider the context of the text being analysed, which aligns with the rationale for removing short or meaningless (to an extent) answers in topic modelling and sentiment analysis. Therefore, it was decided to make a distinction between answers that entail no or a too-short answer (e.g., “Nee” [No]), or answers that did not necessarily answer the question, but did entail indirect meaning (e.g., “Ik heb veel moeite om op de juiste woorden te komen” [I have a lot of trouble finding the right words]). This concluded with a total of 24 responses that were not considered in the analysis, and 16 responses that were saved as a distinct dataset for a possible separate analysis.

To save meaningful data, some responses were duplicated or split to adhere to the two stories better. For example, 18 patients answered both questions in one, leaving the other unanswered. Four other patients summarized their experience, making it fit both questions, allowing for duplication. Additionally, it was necessary to correct spelling mistakes (e.g., “eintje” to “eindje” [a little], “allebij” to “allebei” [both]) and fixing double spacing and symbols (e.g., “ondervindweinig” to “ondervind weinig” [find little], “mn” to “mijn” [my]). On top of that, several patients used certain language or abbreviations to refer to words, which had to be written out (e.g., “zk” to “ziekenhuis” [hospital], “mbv” to “met behulp van” [with the help of], “1ste hulp” to “eerste hulp” [first aid]). Lastly, some patients still used an older version of the Dutch language, which was corrected to the current version (e.g., “nivo” to “niveau” [degree], “kontrole” to “controle” [control]).

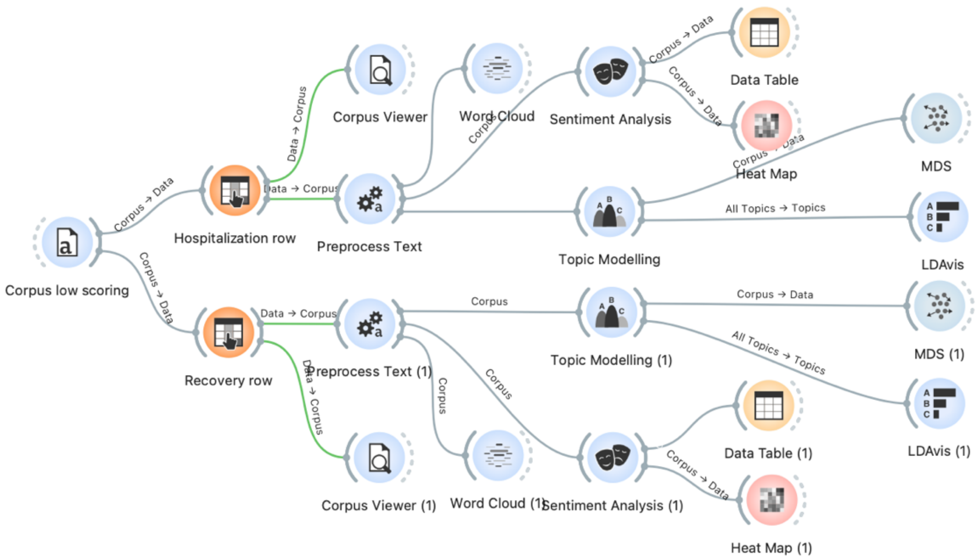
The Orange3 software contains a widget called “Preprocess Text” for pre-processing the data before topic and sentiment analysis. Hence, the text was transformed (conversion to lowercase), as well as filtered using the filtering function, where characters were removed, like commas and periods, and commonly used Dutch stop words were removed (see Appendix B for full list). On top of the exclusion of meaningless words, in the iterative refinement process of topic modelling, words such as “hospital”, “again”, and “oxygen” were found to be pervasive across multiple topics, potentially decreasing topic coherence. To enhance topic specificity and coherence, these ubiquitous terms were selectively excluded from the analysis and included in the list with stop words. Lastly, the tokenization function allowed for the separation of text into words.

2.4.3 Topic modelling

Topic modelling refers to employing methods that uncover the latent semantic structure within texts, offering valuable insights into their diverse themes (Blei, 2012). Topic modelling through machine learning is used, focusing on identifying underlying themes. It focuses on discovering topics using a corpus, through the analysis of clusters of expression in the narrative texts that might be related to each other. For this study, the algorithm of Latent Dirichlet Allocation (LDA) is used; an unsupervised model that can connect similar words, while also distinguishing between words and their meaning. As a result, the output showed a structure containing several topics, with a shared theme for each topic. The following step included the skills of the researcher, to analyse and interpret the themes. The pipeline, including the used widgets, can be seen in Figure 1.

Figure 1.

The Orange3 Pipeline to Topic Modelling and Sentiment Analysis



Topic interpretation in topic modelling can be challenging when relying solely on keywords. This is because common words in the corpus often appear near the top of keyword lists for multiple topics, making it difficult to differentiate their meanings (Sievert & Shirley, 2014). To overcome this, LDAvis, the Orange3 widget called LDAvis was utilized to aid in the interpretability of topics and infer their meaning. It allows visualization of words that occur with high probability in a topic, weighing for words that do not appear frequently in other topics by setting a relevance metric. In this study, a relevance metric of 0.5 was employed to visualize topic keywords that were represented more commonly in the topic than in the general corpus, as advised by Noble and colleagues (2021).

Furthermore, the Orange3 widget Multidimensional Scaling (MDS) was used to create low-dimensional projections of the topics as circles and to fit the distances between these circles (Abayomi-Alli, 2022). The area of each circle (and corresponding values) indicates the Marginal Topic Probability (MTP) of each topic in the corpus (Sievert & Shirley, 2014). Larger circles indicate stronger topic representations, while the proximity of circles indicates shared words among topics and how closely related topics are to each other (Shrader, 2021). In this study, MTP values were utilized to differentiate between small and bigger topics, where a value of 0 means that a topic is not represented at all, and a value of 1 means that only one topic is represented.

The last step included identifying the number of topics used for comparison and interpretation. There is no specific number of topics for a corpus as any topic contains meaning (Weston et al., 2023). Hence, it is the researcher's task to compare topic adherence scores and exclusivity, to indicate how often words within a topic co-occur with other topic's words (Korenčić et al., 2021). The value of a sufficient coherence score is dependent on the data in question and can vary in different settings. However, the aim was to enhance this score. Exclusivity aims to have distinct words in each topic, with the least overlap possible (Weston et al., 2023). Concludingly, the researcher decided to employ 4 topics for the subgroup hospitalization story of patients with low PTSD symptoms and the recovery story of patients with high PTSD symptoms, and 3 topics for the hospitalization story of patients with high PTSD symptoms and the recovery story of patients with low PTSD symptoms.

2.4.4 Sentiment analysis

Sentiment analysis, on the other hand, includes identifying emotions and opinions in a specific text, allowing for an understanding of the expressed sentiments (Medhat et al., 2014), i.e., positive, negative, and neutral. This study applied a lexicon-based approach to identify sentiments in the text. The Orange3 software provided the tools to analyse the text, by using a pre-existing language dataset in Dutch, recognising the words that are presented in the narratives. The valence scores indicate both the sentiment polarity (positive vs negative), as well as its intensity (-1 to 1, 0 indicating neutral). This is done for the two narrative questions separately, so that a distinction can be made between the hospitalization story, and the recovery story, ultimately leading to a more comprehensive understanding of the experience.

The study aims to compare the patients scoring low to patients scoring high on PTSD. Therefore, the differences in the sentiment analysis were first visualized with the heatmap Orange3 widget and analysed in RStudio using the Mann-Whitney U test, code “wilcox.test”.

3. Results

3.1 Topic Modelling

In the topic modelling of hospitalization narratives within the low-scoring PTSD group, four distinct topics were formulated. As can be seen in Table 1, the first topic labelled “An unexpected encounter; initial hospitalization and diagnosis”, predominantly feature keywords such as “moment”, “longen” (lungs), and “bleek” (appeared to be), suggesting descriptions related to the initial phase of hospitalization and diagnostic procedures. For example, one patient wrote the following: “Admitted in May 2021. After a week I thought it would be better but then suddenly, I felt worse. I had to sit on the couch without moving a muscle for three days to save energy and I felt very short of breath. After those days I became very ill with fever.”

Topic two is called “Symptom manifestations and healthcare engagement: how a positive test can spread positivity”, containing keywords like “klachten” (complaints/symptoms), “positief” (positive), and “getest” (tested), indicative of discussions surrounding the presentation of testing upon admission. The word ‘positive’ was used to describe the outcome of the test, but also in terms of sentiment. For example, one patient wrote: “On Monday the 26th of October I received the outcome of the corona test to be positive.”, whereas another patient wrote: “Because my husband and I both have a positive mindset, the hospitalization experience was a smooth process”. More positivity was seen among patients in this subgroup: “I can only be positive about the treatment and the care of the nurses”.

The third topic titled “Navigating the rollercoaster of symptoms and coping at home”, encompasses terms such as “beter” (better), “thuis” (at home), and “voelde” (felt), implying narratives revolving around the progression of symptoms and experiences during homecare. One patient wrote: “After a week I thought that it went better, but then I felt worse again.”. It appeared that the word “home” had several meanings for patients. For example, one patient explained the home to be the place to recover: “At home I was given oxygen for two more weeks.”, where a different patient found it to be more stressful to be back: “It was hard to accept I couldn’t do anything anymore, like working and small jobs at home.”.

Lastly, the fourth topic called “Medical interventions and healing trajectories: negotiating hospital procedures”, included keywords such as “opname” (admission), "longembolie" (pulmonary embolism), and "huisarts" (general practitioner), suggesting hospital procedures and consultations during and after hospitalization. Here, the difference in the experience is observed: "The hospital admission was tough for me", as opposed to: "Oh hospital admission, I was just unlucky because I have a cardiac arrhythmia next to an inflamed pericardium, but it's all over now.".

Concludingly, the topics identified within this subgroup reflect the dimensions of the hospitalization journey, encompassing the initial encounter, the emotions experienced, and the adjustment to the post-hospitalization phase at home. The analysis reveals nuances in the interpretation of words within these topics, where a single term may evoke contrasting sentiments among different patients. This underscores the complexity of attempting to derive meaning from language, particularly among patients with low scores in PTSD.

Table 1

Key Words of each Topic within Low Scoring PTSD Group for the Hospitalization Story with assigned Label and Marginal Topic Probability (MTP)

Topic	Keywords (NL)	Keywords (EN)	Label	MTP
1	Moment, longen, opgenomen, eerste, bleek, huisarts, dag, IC, ziek, erg	Moment, lungs, first, appeared, GP, day, ICU, sick, very	An unexpected encounter; initial hospitalization and diagnosis	0.27
2	Klachten, opgenomen, wel, thuis, positief, getest, november, week, erg, personeel	Complaints, admitted, surely, home, positive, tested, November, week, very, staff	Symptom manifestations and healthcare engagement: how a positive test can spread positivity	0.18
3	Beter, erg, steeds, slechter, thuis, huis, voelde, medicatie, ziek, week	Better, very, still, at home, home, felt, medication, sick, week	Navigating the rollercoaster of symptoms and coping at home	0.30
4	Dag, afdeling, opname, moest, longembolie, huis, opgenomen, gebeld, huisarts, donderdag	Day, department, admission, had to, home, admitted, called, GP, Thursday	Medical interventions and healing trajectories: negotiating hospital procedures	0.24

In the examination of hospitalization narratives among the high-scoring PTSD group, three topics emerged. Table 2 shows the first topic labelled as “Hospitalization environment and personal struggles: solitude and coping with isolation in hospital settings”, predominantly featured keywords like “huis” (house/home), “afdeling” (department), and “kamer” (room), indicating discussions related to the hospital environment and how the patients experienced it. For example: “Fortunately, I was completely done after my hospital admission, the care I received there I could also have gotten at home.”, and “…at this department, there was one nurse who talked to me every time she had to work. She was very sweet, listened to me, and cared. She took the time and didn’t only communicate with me through the intercom. This made the admission more bearable for me, especially because I couldn’t see my loved ones and I was laying there feeling very lonely”.

Topic two, termed “Medication management and treatment procedures with adherence amidst imperatives and the complexity of medication regimes”, encompasses terms such as “medicijnen” (medication), “hart” (heart), and “opgenomen” (admitted), suggesting narratives related to medication and treatment. A patient wrote: “As a result of all the medication and stuff, I also have more complaints regarding my stoma.”, and “Because of corona I’m left falling from one thing into another, my body completely abandoned me.”.

The last topic called “The everlasting result of the traumatic experience while balancing life and responsibilities”, included keywords like “IC” (ICU), “ziek” (sick), and “kinderen” (children), implying the experiences within the ICU and the journey towards recovery post-hospitalization. Additionally, this is the only topic that includes three-time indications, namely “day”, “week”, and “year”, which could mean that the time or duration is important for these patients. For example, “I ended up with a trauma because of my stay in the hospital. I’m still working on processing this daily”, as well as: “Last week it was a year ago that I received oxygen in the ICU. According to the many flashbacks that I have the past weeks because of the hospitalization, I don’t feel like describing it right now.”.

In summary, the identified topics range from feelings of solitude, to adherence to institutional protocols and trauma experiences. Notably, the narratives reveal that for some patients, the traumatic admission hindered their ability to openly discuss their journey. A key point emerges when comparing the low-scoring PTSD group with the high-scoring PTSD group. For instance, while the former group tends to provide a chronological account of their hospital admission journey emphasizing the process from admission to recovery, the latter group delves into their emotional response to the environment. Particularly feelings of isolation and the lasting impact of the admission to their recovery trajectory. Additionally, whereas patients in the low-scoring group appear to exhibit a sense of acceptance regarding their experiences, those with higher PTSD scores tend to linger in the present and past moment.

Table 2

Key Words of each Topic within High Scoring PTSD Group for the Hospitalization Story with assigned Label and Marginal Topic Probability (MTP)

Topic	Keywords (NL)	Keywords (EN)	Label	MTP
1	Huis, afdeling, kamer, man, medicatie, pijn, alleen, opgenomen, bezoek, moest	Home, department, room, man, medication, pain, alone, admitted, visit, had to	Hospitalization environment and personal struggles: solitude and coping with isolation in hospital settings	0.61
2	Moest, echt, medicijnen, thuis, hart, opgenomen, af, artsen, lichaam, tijdens	Had to, really, medicines, at home, heart, admitted, done/off, doctors, body, during	Medication management and treatment procedures with adherence amidst imperatives and the complexity of medication regimes	0.15
3	Week, IC, ziek, zorg, saturatie, dag, opname, fysio, kinderen, jaar	Week, ICU, sick, care, saturation, day, admission, physio, children, year	The everlasting result of the traumatic experience while balancing life and responsibilities	0.23

Table 3 shows the analysis of the recovery narratives among the low-scoring PTSD group. Three topics were formulated, where the first topic is labelled as “Back to life as it was: slow improvements and finding peace by embracing calm and renewal in daily life.” This topic featured keywords such as “last” (burden), “better” (beter), and “voelde” (felt), indicating discussions related to improvements in well-being and overall progress during the recovery process. One patient described the following: “I only have complaints of thinner hair, but that’s already getting better.”.

The second topic called “Return to routine and daily life at home and work” encompasses terms like “wel” (surely), “thuis” (at home), and “werk” (work), suggesting returning to routine activities and daily life post-hospitalization, for example: “I slowly returned to work by doing physiotherapy.”, and “Since September I have been working like normal again, it’s going well!”.

The third topic, titled "Ongoing fatigue management through social support and positivity", included keywords such as "dagen" (days), "moe" (tired), and "gelukkig" (happy/fortunate), implying how patients manage fatigue and the overall recovery process. This is exemplified by the following quote: "…but we went progressively further with our walks, which helped us get back to our old endurance.". Additionally, another patient wrote the following: "Currently I feel just like before the infection. Fortunately. You also hear other stories.".

In conclusion, the recovery stories of those with low scores in PTSD symptomatology, appear to have hope for the future. It was found that these patients have the intrinsic motivation to slowly get back to their old lives. Even though they still experience the effects of the infection, like fatigue, with the support of a significant other and positivity they aim for the best.

Table 3

Key Words of each Topic within Low Scoring PTSD Group for the Recovery with assigned Label and Marginal Topic Probability (MTP)

Topic	Keywords (NL)	Keywords (EN)	Label	MTP
1	Last, beter, voelde, rustig, dagen, fysio, tijd, wel, even, leven, maand	Burden, better, felt, calm, days, physio, time, surely, for a bit, life/living, month	Back to life as it was: slow improvements and finding peace	0.27
2	Wel, goed, thuis, erg, jaar, werk, langzaam, begin, we, beter	Surely, good, at home, very, year, work, slowly, begin, we, better	Return to routine and daily life at home and at work	0.40
3	Dagen, steeds, moe, we, huis, dag, weken, snel, gelukkig, erg	Days, still, tired, we, home, day, weeks, fast, fortunately, very	Ongoing fatigue management through social support and positivity	0.33

Lastly, table 4 shows the recovery narratives among the high-scoring PTSD group, including four distinct topics. The first topic, titled "Daily struggles and recovery efforts while feeling low and lonely", featured keywords such as "vaak" (often), "alleen" (alone), and "pijn" (pain), indicating the daily struggles and efforts exerted during the recovery journey. One quote describes this topic well: "I still sweat a lot; I'm kept awake at night thinking if life is still worth it, but I never had that before Corona. The Corona, and the four hip operations have really had a negative impact on my life.”.

Topic two termed “Experiencing the persistency of symptoms through the slow and long road to recovery” encompassed terms like “oud” (old), “gebruik” (use), and “zeker (certain), suggesting the persistence of symptoms and its influence on recovery. For example: “At the same time it frustrates me that I still don’t have the energy as in my old life.”, as well as “It’s going better little by little, but slowly. The worst complaint is still the fatigue.”.

The third topic, titled “Feeling home at home, moving on from the hospitalization by emotional healing amidst treatment” included keywords like “helaas” (unfortunately), “conditie” (endurance), and “thuiskomst” (homecoming), implying the mental and home-based adjustment as a result from the hospitalization, like “After coming home I was still not able to do anything for a few weeks.”, and “But short of breath I am still, I am still very emotional when I have to write or talk about the past three years. In my sleep, I have seen people that have unfortunately passed away, which also scares me.”.

Finally, topic four was formulated “Reflecting on the experience, moments of relief to lingering memories of struggle”, featuring keywords like “denk” (think), “sinds” (since), and “nooit” (never). These words indicate the reflections patients have on the experience, as well as their mental recovery journey. One patient wrote: “Especially the walking I miss because my feet don’t cooperate. I personally think I must learn to live with it. I sincerely hope that my physical complaints such as rheumatic paints do not get worse.”. Another patient described the following: “I have been back to work for 2-3 mornings a week 2 for several weeks, but due to the back problems, I can’t be as I was before the operations and Corona, and my wish is to be myself again because I haven’t been myself for a long time.”.

In summary, when comparing this subgroup to the others, it stands out that these patients have a more negative view of their recovery than patients scoring low. Not only are more negative words reflected through the topics, like “pain” and “unfortunately”, but the topics also show the drainage and fatigue that won’t improve. For these patients, simply recovering is not easy; they compare themselves to others and see little to no hope left in coping with the persistent symptoms.

Table 4

Key Words of each Topic within High Scoring PTSD Group for the Recovery Story with assigned Label and Marginal Topic Probability (MTP)

Topic	Keywords (NL)	Keywords (EN)	Label	MTP

1	Vaak, leven, moe, pijn, dag, werk, thuis, alleen, lopen, herstel	Often, life/living, tired, pain, day, work, at home, alone, walking, recovery	Daily struggles and recovery efforts while feeling low and lonely	0.61
2	Beter, oude, klachten, vermoeidheid, jaar, langzaam, fysio, gebruik, huis, zeker	Better, old, complaints, fatigue, year, slowly, physio, use, home, certainly	Experiencing the persistency of symptoms through the slow and long road to recovery	0.12
3	Fysio, vanuit, behandeling, we, hersteld, weken, thuis, thuiskomst, conditie, helaas	Physio, from, treatment, we recovered, weeks, at home, homecoming, condition, unfortunately	Feeling home at home, moving on from the hospitalization by emotional healing amidst treatment	0.15
4	Opname, gelukkig, terug, denk, sinds, nooit, moest, tijd, moeten, geheugen	Admission, fortunately, back, think, since, never, had to, time, have to, memory	Reflecting on the experience, moments of relief to lingering memories of struggle	0.12

3.2 Sentiment Analysis

As can be seen in Table 5, the sentiment across subgroups corresponds to varying levels of PTSD scores. Still, distinct patterns emerged. The hospitalization narratives of patients scoring low on PTSD exhibited a slightly more positive sentiment overall (M = 3.83). For example, the sentence “The treatment was good, and the nursing was terrific” received a sentiment score of 5, where “good” and “terrific” are considered 5 and 10 (positive), and “treatment” has a score of -10 (negative). Conversely, the hospitalization narrative of patients scoring high on PTSD displayed a notably more negative sentiment (M = -7.56). Similarly, the recovery narrative of patients scoring low on PTSD demonstrated positive sentiment expressions (M = 13.46), whereas the patients who scored high on PTSD showed a negative sentiment (M = -6.99). An example of scores within this group is the following sentence: “Very hard, experienced many defeats” which was scored with -20, where “very” was given a score of 10, “hard” a 0, “many” a 10, and “defeats” a -40. These mean scores were accompanied by varying standard deviations, ranging from 12.50 to 27.72 across the subgroups, indicating the degree of variability in sentiments.

Table 5

Sentiment Analysis Outcomes of All Subgroups including Mean and Standard Deviation

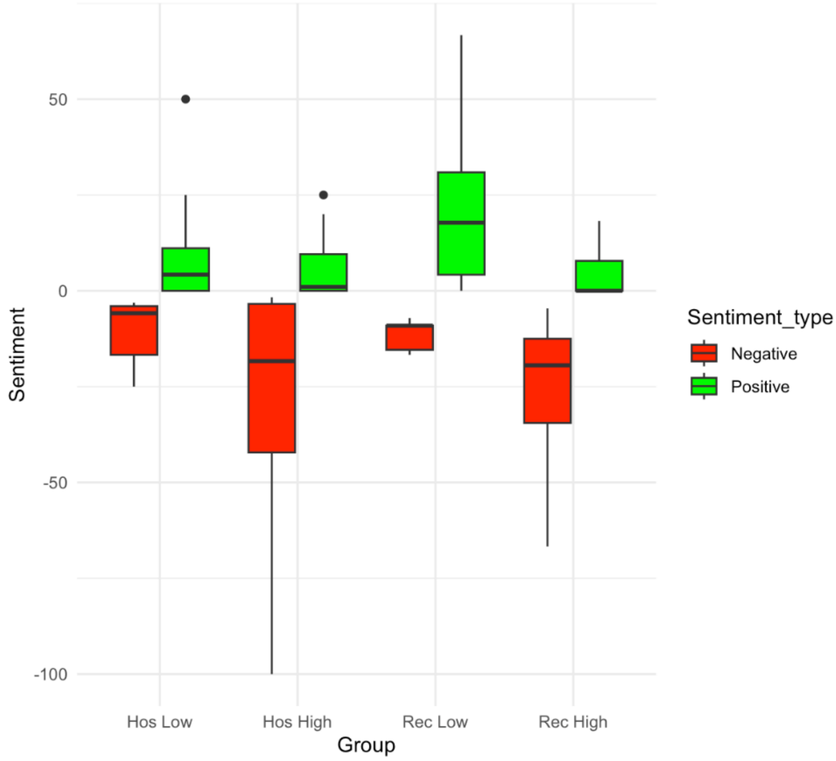
	Hos Low	Hos High	Rec Low	Rec High
Mean	3,83	-7,56	13,46	-6,99
Sd	12,50	27,72	22,80	19,20

Note. Hos = Hospitalization Story, Rec = Recovery Story, Low = Scoring Low on PTSD, High = Scoring High on PTSD, Sd = Standard Deviation

Further exploration revealed nuanced sentiment dynamics within and across subgroups, highlighting the emotional experiences of patients with differing PTSD symptomatologies. The visualization of the distribution is shown in Figure 2. In hospitalization narratives, patients with low PTSD scores exhibit a wider range of sentiments (range: -100 to 50), indicative of diverse emotional experiences during medical treatment. Those with high PTSD scores demonstrate a similar range, but a less extensive positive range (-100 to 25). During the recovery phase, patients with low PTSD scores continue to display a broader range of sentiments (range: -66.7 to 66.7), suggesting ongoing emotional variability. Conversely, patients with high PTSD scores exhibit narrower sentiments during recovery (range: 35.7 to 17.7).

Figure 2

Boxplot Visualization of Sentiment Analysis with Distinction between Positive and Negative Sentiment Distribution across all Subgroups



The findings reveal varied outcomes across the comparisons based on the Mann-Whitney U tests conducted to compare sentiment scores between the groups. For the comparison between sentiment scores of patients with low and high PTSD symptomatology during hospitalization, no statistically significant difference was found ($W = 547$, $p = .14$). Similarly, there was no significant difference in sentiment scores between those in the low and high PTSD groups ($W = 705.50$, $p = .82$). However, a significant difference emerged in the sentiment scores between patients with low and high PTSD symptomatology during the recovery stage ($W = 988.50$, $p <.001$), indicating that sentiment expression varied significantly between these groups. Furthermore, the sentiment analysis highlighted that patients with high PTSD symptomatology tended to exhibit relatively more negative sentiments, whereas those with low PTSD demonstrated relatively more positive sentiments. Additionally, both groups exhibited outliers, contributing to the wide range of sentiment scores that were observed.

4. Discussion

The current study aimed to describe and understand the language use of the patients hospitalized due to COVID-19, and indirectly, their experience of trauma. The methodology involved a mixed-methods approach combining statistical techniques and qualitative interpretation to analyse narrative responses from COVID-19 hospitalized patients.

4.1 Main Findings

Patients scoring low on PTSD exhibit a different thematic focus on their narratives compared to those with high PTSD scores. The hospitalization narratives of low-PTSD patients are characterized by a more procedural view of their medical journey. These patients discussed the timeline of their symptoms and their coping mechanisms both at home and in the hospital. The topics highlight a structured and somewhat accepting view of their experiences. In terms of sentiment, the patients scoring low on PTSD symptomatology exhibit a generally positive sentiment in both stories, often including expressions of gratitude, optimism, and a sense of moving forward, suggesting a more favourable view of their experience, which resonates with research emphasizing the role of resilience in coping with traumatic experiences (Cusack et al., 2022). Moreover, has been found that social support can protect individuals from developing PTSD. Positive social support is associated with being a protective factor against PTSD, reducing the risk of developing the disorder and its impact (Dinenberg et al., 2014). This is carried out by feelings of comfort through the idea of being safe and protected by others which in turn, enhances self-control and emotion regulation.

On the other hand, high-PTSD patients' narratives are characterized by emotional struggle and isolation. Their narratives often reflected the psychological effects of their hospitalization experience. These patients express negative sentiments in both phases. The topics focus on the emotional impact, emphasizing feelings of loneliness and ongoing trauma related to their medical experiences, reflecting consistency with studies highlighting the psychological impact of trauma on emotional well-being. For example, the Social Erosion Model by Perry and colleagues (2023) suggests that PTSD symptoms like avoidance and negative beliefs, could result in a decline in social support. This can potentially increase negative emotions transferred through language use in patients with high PTSD symptomatology. Consequently, certain perceptions of PTSD by society have been linked to reduced help-seeking behaviour, which may further impact the negativity found in the language (Sowers et al., 2022). Cognition also plays a role in PTSD, as negative beliefs about oneself, the world, and the future, can further influence the state of mind of those suffering from PTSD (Je et al., 2022).

When comparing the two stories, hospitalization narratives, regardless of PTSD symptomatology scores, tend to focus on the immediate medical environment, the processes, and the physical symptoms experienced during their stay. For example, patients who score low on PTSD describe detailed procedural elements and their initial coping strategies, while patients who score high on PTSD describe the emotional challenges faced. Research by Abid and colleagues (2023) shows that procedural knowledge, meaning the specific knowledge about procedures in the medical setting (particularly the ICU), plays an important role in the experience of trauma. When patients have a clear understanding of what will be happening and what to expect during their stay, it can potentially diminish their experience of trauma compared to feelings of disorientation. This is exemplified by patients who are well-informed showing reduced anxiety and distress as compared to those who feel uncertain about what will happen to them (Abid et al., 2023). Similarly, literature suggests the crucial role of uncertainty in PTSD, both in general and in specific contexts like the hospital. Traumas involving greater uncertainty have been associated with more severe PTSD symptomatology (Hirsh et al., 2012). Recovery stories, on the other hand, make a shift towards topics of adjustment and ongoing challenges while gaining a positive view of the future, which is in line with the sentiments.

Furthermore, it is important to consider the broader implications of the study in terms of clinical practice, understanding trauma, future pandemics, and technology. Patients with high PTSD symptomatology tend to use more negative words and sentiments in their language, highlighting the importance of addressing emotional and psychological needs in healthcare. For future pandemics or medical situations characterised by significant uncertainty, the study contributes to understanding the need for clear communication in hopes of reducing perceived anxiety and stress. Procedural knowledge may improve the patient's sense of control. Healthcare professionals can be trained to recognize and address the psychological needs of patients, in particular those with high PTSD symptomatology. On top of that, the study supports the implementation of structured support systems, like peer support groups or family-centred care practices to use positive social support as a buffer. Finally, this study sheds light on technological opportunities in trauma. Many patients expressed their gratitude in the narratives for the opportunity to share their story, with several writing multiple paragraphs of text. This underscores the need for research focusing on patient narration. Using technology or more specifically, platforms can provide patients with a way to document and share their experiences or provide psychoeducation about the impact of a hospital admission.

4.2 Strengths and Limitations

One key strength of this study is the extensive design using a mixed-methods approach that combines statistical techniques with qualitative interpretation, which has been found to provide a comprehensive understanding of the topic studied (Johnson & Onwuegbuzie, 2004). The study not only employed comparative analysis but also validated measurement tools like the PCL-5. A second strength is the topic of research, with a focus on bridging the research gap by studying COVID-19 hospitalized patients' experience of trauma. Flores (2005) underscores the impact of medical interpreter services on the quality of healthcare with emphasis on the relevance of addressing language to improve patient care in general, and during global crises. The study provides valuable insights into tailoring care for patients with high PTSD symptomatology, for treatment as well as prevention.

The study also has its limitations. For example, using a cut-off score for the PTSD questionnaire meant that many narratives (full dataset of 432 patients) had to be excluded from the analysis. This limitation is in line with the challenges of sample representativeness and statistical power (Newman et al., 2010), which emphasizes the importance of larger sample sizes for robust research outcomes. Additionally, the use of the PCL-5 as a screening test for PTSD may have limitations, as trauma is complex and can impact the reliability of such tools as explained by Mueser and colleagues (2001). They executed a psychometric evaluation of the assessment, highlighting reliability and validity in measuring trauma and PTSD symptomatology accurately. On top of that, the study's reliance on Orange3 for computational text mining may include flaws that could not have been controlled for. While efforts were made to mitigate these shortcomings through the interpretation of the output by the researcher, some limitations in the software could not be overcome. This is in line with the discussion by Ghazali and Chen (2018) on the reliability and validity of such tools. For example, the sentiment analysis and the score it gives to certain words were at times not in line with the intuition of the researcher. Relying solely on computational tools involves challenges and there will always be a potential for mistakes (Thelwall et al., 2010). Finally, the uncertainty around comorbidities, such as long-COVID and other factors or illnesses, poses a significant limitation in line with the findings of Slover and colleagues (2006). As stated before, the COVID-19 outbreak involves a great deal of uncertainty. Research into long-COVID, like works by Feldman and colleagues (2022) and Orrù and colleagues (2021) have gained some insights into the persistence of symptoms. However, this study did not control for such factors.

4.3 Recommendations for Future Research

First of all, specific PTSD symptoms like avoidance and negative beliefs can be studied through qualitative research, like focus group workshops using thematic analysis to gain deeper personal insights about how these determinants are experienced. A substantial number of patients in this study did not want to answer the question and wrote down being fatigued, or not in the mood to respond. This meant they were excluded from the analysis. However, crucial data was lost, as there is meaning behind these answers opening a door for future research. Dutheil and colleagues (2020) emphasize that individuals with PTSD may avoid seeking help due to factors like stigma, which underscores the importance of understanding these patient perspectives and addressing barriers. Secondly, this study did not control for the comorbidity of other mental illnesses like history of trauma, or substance abuse, despite the likelihood of prevalence (Kang et al., 2019). A study by Banas and Gotwals (2002) has highlighted the significance of comorbid mental health conditions in shaping the experiences of individuals with PTSD. Understanding comorbidities is crucial for comprehensive treatment and recovery. Future research could study language use while controlling for other factors qualitatively and quantitatively. Lastly, while considering the bigger aim to be more prepared for future crises, it is recommended to further develop healthcare strategies to enhance the patient experience through increasing procedural knowledge and support. Here, technological options can be explored. As this study underscored the need for storytelling by patients with low and high PTSD symptomatology during and after hospitalization, future research could explore the use of digital platforms and e-health.

4.4 Conclusion

To conclude, the study provides valuable insights about the experience of COVID-19 patients during and after hospitalization, particularly regarding PTSD symptomatology. The main research question, which focused on understanding the language use of those patients and indirectly their experience of trauma, was answered by identifying differences within the narrative topics and sentiments between patients with low and PTSD symptomatology, as well as the hospitalization and recovery process. Low-PTSD patients tended to describe their experiences with a structured and positive outlook, while high-PTSD patients’ narratives focused on emotional struggles and pessimism.

The mixed methods approach proved to be effective in providing a comprehensive understanding, by shedding light on the importance of addressing the needs of patients and enhancing procedural knowledge to reduce uncertainty. It highlights the role of social support and suggests that clear communication and support systems might be crucial in healthcare settings, especially during global crises. Reflecting on the study’s process, the use of validated tools and computational text mining added accuracy to the analysis. Nevertheless, future research should avoid relying solely on computational analysis, control for comorbidity, and explore specific PTSD-related determinants and symptoms through qualitative methods to gain deeper personal insights. On top of that, it is recommended to explore technological solutions for storytelling and interventions, to further enhance patient care and support. It underscores the need for holistic healthcare approaches addressing both medical and psychological dimensions of care, contributing to a more comprehensive understanding of trauma and resilience in hospitalized patients.

References

Abayomi-Alli, A., Abayomi-Alli, O., Misra, S., & Sanz, L. F. (2022). Study of the Yahoo-Yahoo Hash-Tag tweets using sentiment analysis and opinion mining algorithms. *Information (Basel)*, *13*(3), 152. https://doi.org/10.3390/info13030152

Abid, G., Contreras, F., Rank, S., & Ilyas, S. (2023). Sustainable leadership and wellbeing of healthcare personnel: A sequential mediation model of procedural knowledge and compassion. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1039456

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Atapattu, T., Herath, M., Elvitigala, C., De Zoysa, P., Gunawardana, K., Thilakaratne, M., De Zoysa, K., & Falkner, K. (2022). EmoMent: An Emotion Annotated Mental Health Corpus from two South Asian Countries. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2208.08486

Banas, M., & Gotwals, R. (2022). Modeling the changes in the prevalence of complex PTSD in undergraduate males and females. *Journal of Student Research*, *11*(3). https://doi.org/10.47611/jsrhs.v11i3.2950

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Bo, H., Li, W., Yang, Y., Wang, Y., Zhang, Q., Cheung, T., Wu, X., & Xiang, Y. T. (2020). Posttraumatic stress symptoms and attitude toward crisis mental health services among clinically stable patients with COVID-19 in China. *Psychological Medicine*, *51*(6), 1052–1053. https://doi.org/10.1017/s0033291720000999

Bockhop, F., Zeldovich, M., Cunitz, K., Van Praag, D., Van Der Vlegel, M., Beißbarth, T., Hagmayer, Y., Von Steinbüechel, N., Åkerlund, C., Amrein, K., Anđelić, N., Andreassen, L., Anke, A., Antoni, A., Audibert, G., Azouvi, P., Azzolini, M. L., Bartels, R., Barzó, P., . . . Zoerle, T. (2022). Measurement invariance of six language versions of the post-traumatic stress disorder checklist for DSM-5 in civilians after traumatic brain injury. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-20170-2

Boeschoten, M. A., Bakker, A., Jongedijk, R. A., & Olff, M. (2020). *PTSS checklist voor de DSM-5.* ARQ Nationaal Psychotrauma centrum.

Bury, M. (2001). Illness narratives: fact or fiction? *Sociology of Health and Illness*, *23*(3), 263–285. https://doi.org/10.1111/1467-9566.00252

Chahraoui, K., Laurent, A., Bioy, A., & Quenot, J. (2015). Psychological experience of patients 3 months after a stay in the intensive care unit: A descriptive and qualitative study. *Journal of Critical Care*, *30*(3), 599–605. https://doi.org/10.1016/j.jcrc.2015.02.016

Chamberlain, S. R., Grant, J. E., Trender, W., Hellyer, P. J., & Hampshire, A. (2021). Post-traumatic stress disorder symptoms in COVID-19 survivors: online population survey. *British Journal of Psychiatry Open*, *7*(2). https://doi.org/10.1192/bjo.2021.3

Chi, H., Chiu, N., Chen, C., Weng, S., Lien, C., Lin, C., Hu, Y., Lei, W., Tai, Y., Lin, L., Liu, L. Y., & Lin, C. (2023). To PCR or not? The impact of shifting policy from PCR to rapid antigen tests to diagnose COVID-19 during the omicron epidemic: a nationwide surveillance study. *Frontiers in Public Health*, *11*. https://doi.org/10.3389/fpubh.2023.1148637

Çiftçi, S., Soyluk, Ö., Selek, A., Erol, S., Hekimsoy, Z., Esen, A. D., Dursun, H., Şahin, S., Örük, G., Mert, M., Soylu, H., Yürekli, B. Ş., Ertorer, M. E., Omma, T., Evran, M., Adaş, M., Tanrıkulu, S., Aydín, K., Pekkolay, Z., . . . Yarman, S. (2022). The importance of DHEA-S levels in Cushing's syndrome; Is there a cut-off value in the differential diagnosis? *Hormone and Metabolic Research*, *54*(04), 232–237. https://doi.org/10.1055/a-1783-7901

Cohan, A., Desmet, B., Yates, A., Soldaini, L., MacAvaney, S., & Goharian, N. (2018). SMHD: A Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions. *Computer Science*. https://doi.org/10.48550/arXiv.1806.05258

Cusack, S. E., Bountress, K. E., Denckla, C. A., Group, N. S. F. S. W., Vassileva, J., Dick, D. M., & Amstadter, A. B. (2022). A longitudinal investigation of resilience as a protective factor during the COVID-19 pandemic. *Traumatology*, *28*(3), 403–410. https://doi.org/10.1037/trm0000397

Davydow, D. S., Gifford, J. M., Desai, S. V., Needham, D. M., & Bienvenu, O. J. (2008). Posttraumatic stress disorder in general intensive care unit survivors: a systematic review. *General Hospital Psychiatry*, *30*(5), 421–434. https://doi.org/10.1016/j.genhosppsych.2008.05.006

Dinenberg, R. E., McCaslin, S. E., Bates, M. N., & Cohen, B. E. (2014). Social Support May Protect against Development of Posttraumatic Stress Disorder: Findings from the Heart and Soul Study. *American Journal of Health Promotion*, *28*(5), 294–297. https://doi.org/10.4278/ajhp.121023-quan-511

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infectious Diseases*, *20*(5), 533–534. https://doi.org/10.1016/s1473-3099(20)30120-1

Dutheil, F., Mondillon, L., & Navel, V. (2020). PTSD as the second tsunami of the SARS-Cov-2 pandemic. *Psychological Medicine*, *51*(10), 1773–1774. https://doi.org/10.1017/s0033291720001336

Feldman, D. E., Boudrias, M., & Mazer, B. (2022). Long COVID symptoms in a population-based sample of persons discharged home from hospital. *Canadian Journal of Public Health*, *113*(6), 930–939. https://doi.org/10.17269/s41997-022-00695-9

Ghazali, S. R., & Chen, Y. Y. (2018). Reliability, concurrent validity, and cutoff score of PTSD Checklist (PCL-5) for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition among Malaysian adolescents. *Traumatology*, *24*(4), 280–287. https://doi.org/10.1037/trm0000156

Gray, M. J., & Lombardo, T. W. (2001). Complexity of trauma narratives as an index of fragmented memory in PTSD: a critical analysis. *Applied Cognitive Psychology*, *15*(7). https://doi.org/10.1002/acp.840

Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of Emerging Technologies in Web Intelligence*, *1*(1). https://doi.org/10.4304/jetwi.1.1.60-76

Harrigian, K., Aguirre, C., & Dredze, M. (2020). On the State of Social Media Data for Mental Health Research. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2011.05233

He, Q., Veldkamp, B. P., & De Vries, T. (2012). Screening for posttraumatic stress disorder using verbal features in self narratives: A text mining approach. *Psychiatry Research*, *198*(3), 441–447. https://doi.org/10.1016/j.psychres.2012.01.032

Helms, J., Kremer, S., Merdji, H., Clère-Jehl, R., Schenck, M., Kummerlen, C., Collange, O., Boulay, C., Fafi-Kremer, S., Ohana, M., Anheim, M., & Meziani, F. (2020). Neurologic features in severe SARS-COV-2 infection. *The New England Journal of Medicine*, *382*(23), 2268–2270. https://doi.org/10.1056/nejmc2008597

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychological Review*, *119*(2), 304–320. https://doi.org/10.1037/a0026767

Howes, C., Purver, M., & McCabe, R. (2014). Linguistic Indicators of Severity and Progress in Online Text-based Therapy for Depression. *Association for Computational Linguistics*. https://doi.org/10.3115/v1/w14-3202

Hydén, L. (1997). Illness and narrative. *Sociology of Health and Illness*, *19*(1), 48–69. https://doi.org/10.1111/j.1467-9566.1997.tb00015.x

Je, S., Kim, K., Namgung, S., Lee, S., So, H. S., Choi, J. H., & Choi, H. (2022). A study on the factors affecting anger in patients with post-traumatic stress disorder. *Psychiatry Investigation*, *19*(11), 927–936. https://doi.org/10.30773/pi.2022.0184

Jelinek, L., Stockbauer, C., Randjbar, S., Kellner, M., Ehring, T., & Moritz, S. (2010). Characteristics and organization of the worst moment of trauma memories in posttraumatic stress disorder. *Behaviour Research and Therapy*, *48*(7), 680–685. https://doi.org/10.1016/j.brat.2010.03.014

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods research: a research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26. https://doi.org/10.3102/0013189x033007014

Kaiser, B. N., Varma, S., Carpenter–Song, E., Sareff, R., Rai, S., & Kohrt, B. A. (2020). Eliciting recovery narratives in global mental health: Benefits and potential harms in service user participation. *Psychiatric Rehabilitation Journal*, *43*(2), 111–120. https://doi.org/10.1037/prj0000384

Kang, B., Xu, H., & McConnell, E. S. (2019). Neurocognitive and psychiatric comorbidities of posttraumatic stress disorder among older veterans: A systematic review. *International Journal of Geriatric Psychiatry*, *34*(4), 522–538. https://doi.org/10.1002/gps.5055

Kessler, R. C., Sonnega, A., Bromet, E. J., Hughes, M. L., & Nelson, C. P. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, *52*(12), 1048. https://doi.org/10.1001/archpsyc.1995.03950240066012

Korenčić, D., Ristov, S., Repar, J., & Šnajder, J. (2021). A topic coverage approach to evaluation of topic models. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3109425

Ladds, E., Rushforth, A., Wieringa, S., Taylor, S., Rayner, C. R., Husain, L., & Greenhalgh, T. (2020). Persistent symptoms after Covid-19: qualitative study of 114 "long Covid" patients and draft quality principles for services. *BMC Health Services Research*, *20*(1). https://doi.org/10.1186/s12913-020-06001-y

Ledeneva, Y., & Sidorov, G. (2010). Recent advances in computational linguistics. *Informatica (Lithuanian Academy of Sciences), 34*(1), 3–18. https://dblp.uni-trier.de/db/journals/informaticaSI/informaticaSI34.html#LedenevaS10a

Liu, D., Epskamp, S., Isvoranu, A., Chen, C., Liu, W., & Hong, X. (2021). Network analysis of physical and psychiatric symptoms of hospital discharged patients infected with COVID-19. *Journal of Affective Disorders*, *294*, 707–713. https://doi.org/10.1016/j.jad.2021.07.043

MacAvaney, S., Desmet, B., Cohan, A., Soldaini, L., Yates, A., Zirikly, A., & Goharian, N. (2018). RSDD-Time: Temporal Annotation of Self-Reported Mental Health Diagnoses. *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*. https://doi.org/10.18653/v1/w18-0618

Mao, L., Jin, H., Wang, M., Hu, Y., Chen, S., He, Q., Chang, J., Hong, C., Zhou, Y., Wang, D., Miao, X., Li, Y., & Hu, B. (2020). Neurologic manifestations of hospitalized patients with coronavirus disease 2019 in Wuhan, China. *JAMA Neurology*, *77*(6), 683. https://doi.org/10.1001/jamaneurol.2020.1127

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Mueser, K. T., Rosenberg, S. D., Fox, L., Salyers, M. P., Ford, J. D., & Carty, P. (2001). Psychometric evaluation of trauma and posttraumatic stress disorder assessments in persons with severe mental illness. *Psychological Assessment*, *13*(1), 110–117. https://doi.org/10.1037/1040-3590.13.1.110

Myers, J. R., Solomon, N. P., Lange, R. T., French, L. M., Lippa, S. M., Brickell, T. A., Staines, S., Nelson, J. N., Brungart, D. S., & Coelho, C. (2022). Analysis of discourse production to assess cognitive communication deficits following mild traumatic brain injury with and without posttraumatic stress. *American Journal of Speech-language Pathology*, *31*(1), 84–98. https://doi.org/10.1044/2021_ajslp-20-00281

Newman, M. G., Przeworski, A., Fisher, A. J., & Borkovec, T. D. (2010). Diagnostic comorbidity in adults with Generalized Anxiety Disorder: impact of comorbidity on psychotherapy outcome and impact of psychotherapy on comorbid diagnoses. *Behavior Therapy, 41*(1), 59–72. https://doi.org/10.1016/j.beth.2008.12.005

Nielsen, T. (2017). The stress acceleration hypothesis of nightmares. *Frontiers in Neurology*, *8*. https://doi.org/10.3389/fneur.2017.00201

Noble, P. M., Appleton, C., Radford, A., & Nenadić, G. (2021). Using topic modelling for unsupervised annotation of electronic health records to identify an outbreak of disease in UK dogs. *PloS One*, *16*(12), e0260402. https://doi.org/10.1371/journal.pone.0260402

O'Kearney, R., & Perrott, K. (2006). Trauma narratives in posttraumatic stress disorder: A review. *Journal of Traumatic Stress*, *19*(1), 81–93. https://doi.org/10.1002/jts.20099

Orrù, G., Bertelloni, D., Diolaiuti, F., Mucci, F., Di Giuseppe, M., Biella, M., Gemignani, A., Ciacchini, R., & Conversano, C. (2021). Long-COVID Syndrome? A study on the persistence of neurological, psychological and physiological symptoms. *Healthcare*, *9*(5), 575. https://doi.org/10.3390/healthcare9050575

Papini, S., Yoon, P. J., Rubin, M., López-Castro, T., & Hien, D. A. (2015). Linguistic characteristics in a non-trauma-related narrative task are associated with PTSD diagnosis and symptom severity. *Psychological Trauma: Theory, Research, Practice, and Policy*, *7*(3), 295–302. https://doi.org/10.1037/tra0000019

Paquet, C. B., Cogan, C. M., & Davis, J. L. (2022). Implications of language use in posttraumatic nightmares on psychological symptoms. *Psychological Trauma: Theory, Research, Practice, and Policy*. https://doi.org/10.1037/tra0001266

Peace, K. A., Porter, S., & Brinke, L. T. (2008). Are memories for sexually traumatic events "special"? A within-subjects investigation of trauma and memory in a clinical sample. *Memory*, *16*(1), 10–21. https://doi.org/10.1080/09658210701363583

Penninx, B. W., Benros, M. E., Mann, J. J., & Vinkers, C. H. (2022). How COVID-19 shaped mental health: from infection to pandemic effects. *Nature Medicine*, *28*(10), 2027–2037. https://doi.org/10.1038/s41591-022-02028-2

Perry, N. S., Goetz, D. B., & Shea, M. T. (2023). Longitudinal associations of PTSD and social support by support functions among returning veterans. *Psychological Trauma*, *15*(8), 1346–1354. https://doi.org/10.1037/tra0001190

Ratzer, M., Brink, O., Knudsen, L., & Elklit, A. (2014). Posttraumatic stress in intensive care unit survivors – a prospective study. *Health Psychology and Behavioral Medicine*, *2*(1), 882–898. https://doi.org/10.1080/21642850.2014.943760

Righy, C., Rosa, R. G., Da Silva, R. T. A., Kochhann, R., Migliavaca, C. B., Robinson, C. C., Teche, S. P., Teixeira, C., Bozza, F. A., & Falavigna, M. (2019). Prevalence of posttraumatic stress disorder symptoms in adult critical care survivors: a systematic review and meta-analysis. *Critical Care*, *23*(1). https://doi.org/10.1186/s13054-019-2489-3

Rushforth, A., Ladds, E., Wieringa, S., Taylor, S., Husain, L., & Greenhalgh, T. (2021). Long Covid – The illness narratives. *Social Science & Medicine*, *286*, 114326. https://doi.org/10.1016/j.socscimed.2021.114326

Sandman, N., Valli, K., Kronholm, E., Vartiainen, E., Laatikainen, T., & Paunio, T. (2017). Nightmares as predictors of suicide: an extension study including war veterans. *Scientific Reports*, *7*(1). https://doi.org/10.1038/srep44756

Sawalha, J., Yousefnezhad, M., Shah, Z., Brown, M., Greenshaw, A. J., & Greiner, R. (2022). Detecting presence of PTSD using sentiment analysis from text data. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.811392

Seo, J., Kim, S. E., Kim, Y., Kim, E. J., Kim, T., Kim, T. H., Lee, S. H., Lee, E., Lee, J., Seo, Y. B., Jeong, Y., Jung, Y. H., Choi, Y. J., & Song, J. Y. (2024). Updated Clinical Practice Guidelines for the Diagnosis and Management of long COVID. *Infection And Chemotherapy/Gam'yeom Gwa Hwahag Yo'beob/Infection & Chemotherapy*, *56*(1), 122. https://doi.org/10.3947/ic.2024.0024

Shih, R. A., Schell, T. L., Hambarsoomian, K., Belzberg, H., & Marshall, G. N. (2010). Prevalence of posttraumatic stress disorder and major depression after trauma center hospitalization. *Journal of Trauma-injury Infection and Critical Care*, *69*(6), 1560–1566. https://doi.org/10.1097/ta.0b013e3181e59c05

Short, J. C., McKenny, A. F., & Reid, S. W. (2018). More than words? Computer-Aided text analysis in organizational behavior and psychology research. *Annual Review of Organizational Psychology and Organizational Behavior*, *5*(1), 415–435. https://doi.org/10.1146/annurev-orgpsych-032117-104622

Shrader, C. B., Ravenscroft, S., Kaufmann, J. B., & Hansen, K. (2020). Collusion among Accounting Students: Data Visualization and Topic Modeling of Student Interviews. *Decision Sciences Journal of Innovative Education*, *19*(1), 40–62. https://doi.org/10.1111/dsji.12226

Sievert, C., & Shirley, K. E. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. https://doi.org/10.3115/v1/w14-3110

Slover, J., Abdu, W. A., Hanscom, B., & Weinstein, J. N. (2006). The impact of comorbidities on the change in Short-Form 36 and Oswestry scores following lumbar spine surgery. *Spine (Philadelphia, Pa. 1976)/Spine*, *31*(17), 1974–1980. https://doi.org/10.1097/01.brs.0000229252.30903.b9

Sowers, A. F., Clapp, J. D., Freng, S. A., & Stanton, K. (2023). The folk psychiatry model: Developing a measure of lay perceptions of PTSD. Psychological Assessment, 35(12), 1120–1133. https://doi.org/10.1037/pas0001274

Stanley, I. H., Hom, M. A., Spencer-Thomas, S., & Joiner, T. E. (2017). Examining anxiety sensitivity as a mediator of the association between PTSD symptoms and suicide risk among women firefighters. *Journal of Anxiety Disorders*, *50*, 94–102. https://doi.org/10.1016/j.janxdis.2017.06.003

Tedstone, J. E., & Tarrier, N. (2003). Posttraumatic stress disorder following medical illness and treatment. *Clinical Psychology Review*, *23*(3), 409–448. https://doi.org/10.1016/s0272-7358(03)00031-x

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*(12), 2544–2558. https://doi.org/10.1002/asi.21416

Thornton, E., Patalay, P., Matthews, D., & Bannard, C. (2021). Does early child language predict internalizing symptoms in adolescence? An investigation in two birth cohorts born 30 years apart. *Child Development*, *92*(5), 2106–2127. https://doi.org/10.1111/cdev.13615

Tohme, P., Grey, I., El-Tawil, M. T., Maouch, M. E., & Abi-Habib, R. (2024). Prevalence and correlates of mental health difficulties following the beirut port explosion: The roles of mentalizing and resilience. *Psychological Trauma: Theory, Research, Practice, and Policy*, *16*(1), 30–38. https://doi.org/10.1037/tra0001328

Trusko, B., Rosenbloom, S. T., Montella, D., Jackson, J. C., FitzHenry, F., Brown, S. H., Elkin, P. L., Fielstein, E. M., Kotter, K., Tuttle, M. S., Iannelli, R. J., & Speroff, T. (2010). Are posttraumatic stress disorder mental health terms found in SNOMED-CT medical terminology. *Journal of Traumatic Stress*, *23*(6), 794–801. https://doi.org/10.1002/jts.20591

Van Praag, D., Fardzadeh, H. E., Čović, A., Maas, A. I. R., & Von Steinbüchel, N. (2020). Preliminary validation of the Dutch version of the Posttraumatic stress disorder checklist for DSM-5 (PCL-5) after traumatic brain injury in a civilian population. *PLOS ONE*, *15*(4), e0231857. https://doi.org/10.1371/journal.pone.0231857

Vanderlind, W. M., Rabinovitz, B. B., Miao, I. Y., Oberlin, L. E., Bueno-Castellano, C., Fridman, C. B., Jaywant, A., & Kanellopoulos, D. (2021). A systematic review of neuropsychological and psychiatric sequalae of COVID-19: implications for treatment. *Current Opinion in Psychiatry*, *34*(4), 420–433. https://doi.org/10.1097/yco.0000000000000713

Vindegaard, N., & Benros, M. E. (2020). COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. *Brain Behavior and Immunity*, *89*, 531–542. https://doi.org/10.1016/j.bbi.2020.05.048

Weston, S. J., Shryock, I., Light, R., & Fisher, P. A. (2023). Selecting the Number and Labels of Topics in Topic Modeling: A Tutorial. *Advances in Methods and Practices in Psychological*

Science, *6*(2). https://doi.org/10.1177/25152459231160105

WHO Coronavirus (COVID-19) dashboard. (n.d.). WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. https://covid19.who.int/

Xiang, Y. T., Yang, Y., Li, W., Zhang, L., Zhang, Q., Cheung, T., & Ng, C. H. (2020). Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. *The Lancet Psychiatry*, *7*(3), 228–229. https://doi.org/10.1016/s2215-0366(20)30046-8

Xiao, S., Luo, D., & Yang, X. (2020). Survivors of COVID-19 are at high risk of posttraumatic stress disorder. *Global Health Research and Policy*, *5*(1). https://doi.org/10.1186/s41256-020-00155-2

Zatzick, D., Marmar, C. R., Weiss, D. S., Browner, W. S., Metzler, T. J., Golding, J. M., Stewart, A. L., Schlenger, W. E., & Wells, K. B. (1997). Posttraumatic stress Disorder and functioning and quality of life outcomes in a nationally representative sample of male Vietnam veterans. *American Journal of Psychiatry*, *154*(12), 1690–1695. https://doi.org/10.1176/ajp.154.12.1690

Zatzick, D., Rivara, F. P., Nathens, A. B., Jurkovich, G. J., Wang, J., Fan, M., Russo, J., Salkever, D. S., & MacKenzie, E. J. (2007). A nationwide US study of post-traumatic stress after hospitalization for physical injury. *Psychological Medicine*, *37*(10), 1469–1480. https://doi.org/10.1017/s0033291707000943

Zhang, J., Lu, H., Zeng, H., Zhang, S., Du, Q., Jiang, T., & Du, B. (2020). The differential psychological distress of populations affected by the COVID-19 pandemic. *Brain Behavior and Immunity*, *87*, 49–50. https://doi.org/10.1016/j.bbi.2020.04.031

Appendix A

RStudio Syntax for Data Pre-processing and Comparative Analysis

```
install.packages('tidyverse')
install.packages("tidyr")
install.packages("readxl")
install.packages("dplyr")
install.packages("writexl")
library(readxl)
library(tidyverse)
library(tidyr)
library(dplyr)
library(writexl)

Nar_PCL_6mnd_new = read_xlsx('6mnd_nar+PCL5_zondertwijfelachtigevragen.xlsx', .name_repair
= "unique_quiet")
NarPCL_PosGez_6mnd_new =
read_xlsx('6mnd_nar+PCL5+posgez_zonder.twijfelachtigevragenxlsx.xlsx', .name_repair =
"unique_quiet")
Nar_PCL_12mnd_new = read_xlsx('12mnd_nar+PCL5_zondertwijfelachtigevragen.xlsx',
.name_repair = "unique_quiet")

# Change names of narration questions to ...stories
Nar_PCL_6mnd_new <- Nar_PCL_6mnd_new %>%
  rename(hospitalization_story = Q24,
         recovery_story = Q25)
NarPCL_PosGez_6mnd_new <- NarPCL_PosGez_6mnd_new %>%
  rename(hospitalization_story = Q70,
         recovery_story = Q71)
Nar_PCL_12mnd_new <- Nar_PCL_12mnd_new %>%
  rename(hospitalization_story = Q33,
         recovery_story = Q32)

# Change column name Q1_1 to ResponseID
Nar_PCL_6mnd_new <- Nar_PCL_6mnd_new %>%
  rename(ResponseID = Q1_1)
NarPCL_PosGez_6mnd_new <- NarPCL_PosGez_6mnd_new %>%
  rename(ResponseID = Q1_1)
Nar_PCL_12mnd_new <- Nar_PCL_12mnd_new %>%
  rename(ResponseID = Q1_1)

view(NarPCL_PosGez_6mnd_new)
view(Nar_PCL_6mnd_new)
view(Nar_PCL_12mnd_new)

# Selecting only the relevant columns (stories & PCL-5 questionnaire)
Nar_PCL_6mnd_new <- Nar_PCL_6mnd_new %>%
  select(ResponseID, hospitalization_story, recovery_story, Q26_1, Q26_2, Q26_3, Q26_4, Q26_5,
Q26_6, Q26_7, Q26_8, Q26_9, Q26_10,
         Q26_11, Q26_12, Q26_13, Q26_14, Q26_15, Q26_16, Q26_17, Q26_18, Q26_19, Q26_20)
NarPCL_PosGez_6mnd_new <- NarPCL_PosGez_6mnd_new %>%
  select(ResponseID, hospitalization_story, recovery_story, Q72_1, Q72_2, Q72_3, Q72_4, Q72_5,
Q72_6, Q72_7, Q72_8, Q72_9, Q72_10,
         Q72_11, Q72_12, Q72_13, Q72_14, Q72_15, Q72_16, Q72_17, Q72_18, Q72_19, Q72_20)
Nar_PCL_12mnd_new <- Nar_PCL_12mnd_new %>%
```

```
  select(ResponseID, hospitalization_story, recovery_story, Q31_1, Q31_2, Q31_3, Q31_4, Q31_5,
Q31_6, Q31_7, Q31_8, Q31_9, Q31_10,
        Q31_11, Q31_12, Q31_13, Q31_14, Q31_15, Q31_16, Q31_17, Q31_18, Q31_19, Q31_20)

# Add a new column 'Total_Score' with the sum of responses for each row - !!!! Nar_PCL_6mnd !!!!

question_columns <- c("Q26_1", "Q26_2", "Q26_3", "Q26_4", "Q26_5", "Q26_6", "Q26_7",
"Q26_8", "Q26_9",
                 "Q26_10", "Q26_11", "Q26_12", "Q26_13", "Q26_14", "Q26_15", "Q26_16",
"Q26_17",
                 "Q26_18", "Q26_19", "Q26_20")

Nar_PCL_6mnd_new <- Nar_PCL_6mnd_new %>%
  mutate(across(all_of(question_columns), ~ recode(.,
                                          "Helemaal niet" = 0,
                                          "Een beetje" = 1,
                                          "Matig" = 2,
                                          "Nogal veel" = 3,
                                          "Extreem veel" = 4,
                                          .default = -1)))

Nar_PCL_6mnd_new$Total_Score <- rowSums(Nar_PCL_6mnd_new[, question_columns], na.rm =
TRUE)

# Add a new column 'Total_Score' with the sum of responses for each row - !!!!
NarPCL_PosGez_6mnd !!!!

question_columns <- c("Q72_1", "Q72_2", "Q72_3", "Q72_4", "Q72_5", "Q72_6", "Q72_7",
"Q72_8", "Q72_9", "Q72_10",
                 "Q72_11", "Q72_12", "Q72_13", "Q72_14", "Q72_15", "Q72_16", "Q72_17",
"Q72_18", "Q72_19", "Q72_20")

NarPCL_PosGez_6mnd_new <- NarPCL_PosGez_6mnd_new %>%
  mutate(across(all_of(question_columns), ~ recode(.,
                                          "Helemaal niet" = 0,
                                          "Een beetje" = 1,
                                          "Matig" = 2,
                                          "Nogal veel" = 3,
                                          "Extreem veel" = 4,
                                          .default = -1)))

NarPCL_PosGez_6mnd_new$Total_Score <- rowSums(NarPCL_PosGez_6mnd_new[,
question_columns], na.rm = TRUE)

# Add a new column 'Total_Score' with the sum of responses for each row - !!!! Nar_PCL_12mnd !!!!

question_columns <- c("Q31_1", "Q31_2", "Q31_3", "Q31_4", "Q31_5", "Q31_6", "Q31_7",
"Q31_8", "Q31_9",
                 "Q31_10", "Q31_11", "Q31_12", "Q31_13", "Q31_14", "Q31_15", "Q31_16",
"Q31_17",
                 "Q31_18", "Q31_19", "Q31_20")

Nar_PCL_12mnd_new <- Nar_PCL_12mnd_new %>%
  mutate(across(all_of(question_columns), ~ recode(.,
                                          "Helemaal niet" = 0,
```

```
                                          "Een beetje" = 1,
                                          "Matig" = 2,
                                          "Nogal veel" = 3,
                                          "Extreem veel" = 4,
                                          .default = -1)))

Nar_PCL_12mnd_new$Total_Score <- rowSums(Nar_PCL_12mnd_new[, question_columns], na.rm
= TRUE)

# Now select only what's necessary
Nar_PCL_6mnd_new <- Nar_PCL_6mnd_new %>%
  select(ResponseID, hospitalization_story, recovery_story, Total_Score)
NarPCL_PosGez_6mnd_new <- NarPCL_PosGez_6mnd_new %>%
  select(ResponseID, hospitalization_story, recovery_story, Total_Score)
Nar_PCL_12mnd_new <- Nar_PCL_12mnd_new %>%
  select(ResponseID, hospitalization_story, recovery_story, Total_Score)

# Merge datasets together, based on ResponseID
LTI_Demo_Complete <- read_sav("LTI Demo Complete.sav")

merged_dataset <- bind_rows(Nar_PCL_6mnd_new, NarPCL_PosGez_6mnd_new,
Nar_PCL_12mnd_new)

LTI_Demo_Complete <- LTI_Demo_Complete %>%
  rename(ResponseID = ID)

columns_to_include <- c("ResponseID", "Age", "Gender", "Weight_KG", "Condition_Depression",
"AlcoholFreq")

Merged_NarPCLDemogr <- merge(merged_dataset, LTI_Demo_Complete[columns_to_include], by =
"ResponseID", all.x = TRUE)

# delete some NA's specifically
# 1
rows_to_delete <- c(1:45)

Merged_NarPCLDemogr <- Merged_NarPCLDemogr %>%
  slice(-rows_to_delete)

rows_to_delete <- c(35, 37, 38, 41, 44, 52, 53)

Merged_NarPCLDemogr <- Merged_NarPCLDemogr %>%
  slice(-rows_to_delete)

# Count how many patients have a score of 28 or higher
num_high_scoring_participants <- sum(Merged_NarPCLDemogr$Total_Score >= 28, na.rm = TRUE)
print(num_high_scoring_participants)

# Count participants with Total_Score = 0
num_zero_score_participants <- sum(Merged_NarPCLDemogr$Total_Score == 0, na.rm = TRUE)
print(num_zero_score_participants)

# Count participants with Total_Score between 1 and 27
num_non_zero_score_participants <- sum(Merged_NarPCLDemogr$Total_Score > 0 &
                                  Merged_NarPCLDemogr$Total_Score <= 27, na.rm = TRUE)
```

```
print(num_non_zero_score_participants)

# Mean Age of high scoring participants
mean(high_scoring_participants$Age, na.rm = TRUE)
sd(high_scoring_participants$Age, na.rm = TRUE)
table(high_scoring_participants$Gender)

# Filter participants with Total_Score from 1 to 5
participants_1_to_5 <- Merged_NarPCLDemogr %>%
  filter(Total_Score >= 1 & Total_Score <= 5)

View(Merged_NarPCLDemogr)
# make a low scoring group with similar demographics as the high scoring group

mean_age_high_scoring <- mean(high_scoring_participants$Age, na.rm = TRUE)
high_scoring_participants$Gender <- as.factor(high_scoring_participants$Gender)
num_males_high_scoring <- sum(high_scoring_participants$Gender == "1", na.rm = TRUE)
num_females_high_scoring <- sum(high_scoring_participants$Gender == "2", na.rm = TRUE)

# Step 2: Filter "participants_1_to_5" Dataset
participants_1_to_5_filtered <- participants_1_to_5 %>%
  filter(Total_Score >= 1 & Total_Score <= 5 & !is.na(Gender))

# Step 3: Sample Participants with Similar Demographics
low_scoring_dataset <- participants_1_to_5_filtered %>%
  group_by(Gender) %>%
  sample_n(num_males_high_scoring + num_females_high_scoring, replace = TRUE) %>%
  ungroup() %>%
  filter(abs(mean(Age, na.rm = TRUE) - mean_age_high_scoring) <= 5)  # Allow a difference of up to
5 years in mean age

low_scoring_dataset <- low_scoring_dataset %>%
  sample_n(35)

#check demographics
mean(low_scoring_dataset$Age, na.rm = TRUE)
sd(low_scoring_dataset$Age, na.rm = TRUE)
table(low_scoring_dataset$Gender)

mean(low_scoring_dataset$Total_Score, na.rm = TRUE)
sd(low_scoring_dataset$Total_Score, na.rm = TRUE)

#check demographics
mean(high_scoring_participants$Age, na.rm = TRUE)
sd(high_scoring_participants$Age, na.rm = TRUE)
table(high_scoring_participants$Gender)

mean(high_scoring_participants$Total_Score, na.rm = TRUE)
sd(high_scoring_participants$Total_Score, na.rm = TRUE)

# Save low scoring participants dataset with row names as the first row
write_xlsx(low_scoring_dataset, "low_scoring_dataset", row.names = TRUE)
# Voeg de rijnamen toe als eerste rij aan je dataset
low_scoring_dataset_with_rownames <- cbind("RowNames" = rownames(low_scoring_dataset),
low_scoring_dataset)
```

```
# Sla de dataset op als een Excel-bestand zonder de rijnamen te gebruiken
write_xlsx(low_scoring_dataset, "low_scoring_dataset")

# Save high scoring participants dataset with row names as the first row
write_xlsx(high_scoring_participants, "high_scoring_participants", row.names = TRUE)

topic_sentiment_data <- data.frame(
  Topic = c("HOS LOW", "HOS HIGH", "REC LOW", "REC HIGH"),
  Negative = c(-1, -2, -3, -4),
  Neutral = c(0, 0, 0, 0),
  Positive = c(1, 2, 3, 4))

contingency_table <- as.table(frequencies)
rownames(contingency_table) <- topics
colnames(contingency_table) <- sentiments

# Step 3: Perform Fisher's exact test
fisher_result <- fisher.test(contingency_table)

# View the test result
print(fisher_result)
# Print the data frame
print(topic_sentiment_data)

# Contingency table for HOS LOW subgroup
hos_low <- data.frame(
  Topic = c(1, 2, 3, 4),
  Sentiment = c(0, 10.5, 0, 0),
  stringsAsFactors = FALSE)

# Contingency table for HOS HIGH subgroup
hos_high <- data.frame(
  Topic = c(1, 2, 3),
  Sentiment = c(-10.5, 5.3, -5.3),
  stringsAsFactors = FALSE)

# Contingency table for REC LOW subgroup
rec_low <- data.frame(
  Topic = c(1, 2, 3),
  Sentiment = c(9.5, 15.8, 5.3),
  stringsAsFactors = FALSE)

# Contingency table for REC HIGH subgroup
rec_high <- data.frame(
  Topic = c(1, 2, 3, 4),
  Sentiment = c(-10.5, -10.5, -5.3, 0),
  stringsAsFactors = FALSE)

# Calculate observed frequencies for each subgroup
hos_low_table <- table(hos_low$Topic, hos_low$Sentiment)
hos_high_table <- table(hos_high$Topic, hos_high$Sentiment)
rec_low_table <- table(rec_low$Topic, rec_low$Sentiment)
rec_high_table <- table(rec_high$Topic, rec_high$Sentiment)

# Conduct chi-square test for each subgroup
```

```
hos_low_chisq <- chisq.test(hos_low_table)
hos_high_chisq <- chisq.test(hos_high_table)
rec_low_chisq <- chisq.test(rec_low_table)
rec_high_chisq <- chisq.test(rec_high_table)

# Print the chi-square test results
print("HOS LOW Chi-Square Test:")
print(hos_low_chisq)
print("HOS HIGH Chi-Square Test:")
print(hos_high_chisq)
print("REC LOW Chi-Square Test:")
print(rec_low_chisq)
print("REC HIGH Chi-Square Test:")
print(rec_high_chisq)
```

Appendix B

List of Dutch Stop Words Excluded from Analysis

Ehm	zich zij	maakt
Uh	zijn	weet
Eigenlijk	zo	weten
Dus	zoals	mocht
Nou	zodat	werden
Zeg maar	gewoon	mochten
weet	Maar	komen
Enzovoort	Kijk	komt
Enzovoorts	krijgen	kwam
Enzo	daarna	heel
En zo	nadat	elke
Graag	verder	per
Eigenlijk	kreeg	keer
Echter	daarna	weer
Om eerlijker te zijn	daarvoor	doordat
In principe	gaan	door
Juist	gehad	ziekenhuis
Zoals u weet	zeggen	corona
Bijvoorbeeld	waardoor	covid
Op die manier	gegaan	zuurstof
Joh	ging	steeds
Sowieso	maken	hele
Jawel	daarnaast	eerste
A b c d e f g h i j k l m n o p q r s t u v w x y z	maak	kreeg