



# Text mining to improve screening for trauma-related symptoms in a global sample

D. Marengo<sup>a,1</sup>, C.M. Hoeboer<sup>b,1,\*</sup>, B.P. Veldkamp<sup>c</sup>, GPS-txt consortium<sup>#</sup>, M. Olf<sup>b,d</sup>

<sup>a</sup> Department of Psychology, University of Turin, Via Verdi 10, Turin 10124, Italy

<sup>b</sup> Department of Psychiatry, Amsterdam Public Health, University of Amsterdam, Amsterdam UMC location, Meibergdreef 9, Amsterdam 1105 AZ, the Netherlands

<sup>c</sup> ARQ National Psychotrauma Centre, Diemen, the Netherlands

<sup>d</sup> Department of Learning, Data-Analytics, and Technology, Faculty of Behavioral Management and Social Sciences, University of Twente, the Netherlands

## ARTICLE INFO

### Keywords:

Trauma-related symptoms  
Text mining  
PTSD  
Screening

## ABSTRACT

Previous studies showed that textual information could be used to screen respondents for posttraumatic stress disorder (PTSD). In this study, we explored the feasibility of using language features extracted from short text descriptions respondents provided of stressful events to predict trauma-related symptoms assessed using the Global Psychotrauma Screen. Texts were analyzed with both closed- and open-vocabulary methods to extract language features representing the occurrence of words, phrases, or specific topics in the description of stressful events. We also evaluated whether combining language features with self-report information, including respondents' demographics, event characteristics, and risk factors for trauma-related disorders, would improve the prediction performance. Data were collected using an online survey on a cross-national sample of 5048 respondents. Results showed that language data achieved the highest predictive power when both closed- and open-vocabulary features were included as predictors. Combining language data and self-report information resulted in a significant increase in performance and in a model which achieved good accuracy as a screener for probable PTSD diagnosis ( $.7 < \text{AUC} \leq .8$ ), with similar results regardless of the length of the text description of the event. Overall, results indicated that short texts add to the detection of trauma-related symptoms and probable PTSD diagnosis.

## 1. Introduction

After the experiencing a stressful event such as a serious injury, sexual violence, or a life-threatening situation including an intensive care admission due to the Coronavirus disease-2019 (COVID-19), some people develop persistent mental health problems. Common problems include posttraumatic stress disorder (PTSD; de Vries and Olf, 2009), depression (Breslau et al., 2000), and sleeping problems (Milanek et al., 2019). Trauma-related disorders tend to persist for years when untreated (Kessler et al., 2017) and are related to many adverse outcomes such as unemployment, reduced quality of life, and suicidal behavior (Kessler et al., 2010, Pagotto et al., 2015). Trauma and trauma-related disorders also affect long-term physical health including cardiovascular disease and obesity (Kuhar and Kocjan, 2021). When accurately

identified, many trauma-related disorders are effectively treatable with psychotherapy within a short timespan (Cuijpers et al., 2013, Mavranzouli et al., 2020, Trauer et al., 2015). Numerous studies have shown that the recent COVID-19 pandemic is related to increased prevalence rates of trauma-related disorders in healthcare workers (Marvaldi et al., 2021) and the general population (Arora et al., 2020). Since the pandemic (in)directly affects a large proportion of the population, many are at risk for developing trauma-related disorders. Hence, efficiently and accurately identifying those who (are likely to) suffer from trauma-related disorders is crucial to provide treatment and prevent other adverse outcomes. Considering trauma as a public health issue (Denckla et al., 2020, Magruder et al., 2017) emphasizes the need for accurate and easy screening.

Short screening instruments have been developed, such as the Global

\* Corresponding author.

E-mail address: [c.m.hoeboer@amsterdamumc.nl](mailto:c.m.hoeboer@amsterdamumc.nl) (C.M. Hoeboer).

<sup>1</sup> Marengo, D. and Hoeboer, C.M. share first authorship.

<sup>#</sup> GPS-txt consortium: M. Oe, A. Bakker, A. Hovnanyan, E. Bröcker, X. Anastassiou-Hadjicharalambous, B.M. Coimbra, H. Aakvaag, J.D. Javakhishvili, M.H. Ozgen, L. Sales, D. Rabellino, B. Lueger-Schuster, Y. Qing, E. Grace, I. Primasari, M. Dragan, C. Salgado, D. Ajdukovic.

<https://doi.org/10.1016/j.psychres.2022.114753>

Received 2 February 2022; Received in revised form 22 July 2022; Accepted 27 July 2022

Available online 28 July 2022

0165-1781/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Psychotrauma Screen (Olf et al., 2020), which screens for many trauma-related disorders and provides a total trauma-related symptom score. Although the GPS is more efficient than a clinician-administered interview, developing even shorter screening instruments that do not require symptom inventarisation might increase the screening efficiency and lower the burden for respondents, especially when confronted with situations such as COVID-19. One possibility to improve the screening process for trauma-related disorders that showed promise in previous studies is the use of text mining techniques (He et al., 2012, 2019). The idea of these techniques is to use textual information respondents provide to predict their trauma-related symptoms and disorders. Textual information provided by respondents can be processed to extract language features, which may consist of the proportion of words referring to specific topics naturally emerging from the data, count of specific words (or combinations of consecutive words, i.e., n-grams), as well as features reflecting affective, social, cognitive, and perceptual processes as obtained using existing lexicons (e.g., the LIWC lexicon; Pennebaker et al., 2007). By design, many questionnaires assessing trauma-related disorders already include an open-ended question that asks participants to describe the stressful event currently affecting them the most. Usually, this question is only used to establish whether the stressful event meets the 'A-criterion' for PTSD (APA, 2013). However, the language respondents use to describe the stressful experience may provide essential information predictive of their symptoms. For instance, many trauma-related symptoms directly refer back to this event (e.g., avoidance of triggers related to the event or emotional reactions related to reminders of the event). These symptoms might therefore be reflected in the response of participants when asked to describe this event (APA, 2013).

Previous studies have shown that textual information from self-statements can be used to screen for trauma-related disorders such as PTSD (He et al., 2012, 2019). When combined with a few other relevant predictors of trauma-related symptoms such as item-based measures, this method might lead to efficient screening methods (He et al., 2019). In the current study, we use this method in a novel way by applying it to the text respondents used to describe the stressful event they experienced.

### 1.1. Current study

In the present study, we sought to determine the relevance of a set of language features extracted from textual descriptions of stressful events to predict trauma-related symptoms in a large global population. We also compared the relevance of language features for the purpose of prediction with that of self-reported demographics, event characteristics, and risk factors for trauma-related symptoms. Textual and self-reported information were assessed by administering an online survey including a validated instrument screening for a range of trauma-related psychological problems, namely the Global Psychotrauma Screen (GPS symptom score; Olf et al., 2020). We predicted trauma-related symptoms using a machine learning approach, consisting of the cross-validation of predictive models performed using three different algorithms: Elastic-Net, Random Forest, and a stacking ensemble of these algorithms (i.e., we use a "super learner" approach; van der Laan et al., 2007). We chose a machine learning approach instead of a standard parametric method because machine learning methods can help to identify hidden interactions and non-linearity among features in predicting the outcome, and because they can help reduce the risk of overestimation prediction performance (i.e., overfitting; for a review of application of machine-learning in studying PTSD and other trauma-related disorders, see Ramos-Lima et al., 2020) Using these methods, we examined the predictive power of the following set of features, either alone and or in combination with each other: (1) language features extracted from the open-ended section of the screener, (2) demographics, (3) event characteristics and (4) risk factors for trauma-related symptoms. Finally, we evaluated the contribution of

these features for the purpose of creating a screening procedure for detecting probable PTSD diagnosis, assessed by splitting the sample according to an existing cut-off for PTSD. We also checked the role of the length of the description of the events provided by participants in improving or worsening the prediction accuracy.

## 2. Method

### 2.1. Sample and procedure

We used data from the Global Psychotrauma Screen – Cross Cultural responses to COVID-19 study (GPS-CCC; [www.global-psychotrauma.net/covid-19-projects](http://www.global-psychotrauma.net/covid-19-projects)). The main aim of the GPS-CCC study was to compare COVID-19-related stressful events with other types of stressful events around the globe (see main publication by Olf et al., 2021). For this study, 6114 participants from 88 countries were recruited during the COVID-19 pandemic between the 21st of March and the 3rd of September in 2020 through online advertisements from members of the Global Collaboration on Traumatic Stress ([www.global-psychotrauma.net](http://www.global-psychotrauma.net)). Participants were included in the present study if they were 16 years old or older ( $n$  excluded = 16) and if they provided an interpretable written text ( $n$  excluded = 1050), resulting in a final sample of 5048 participants (Mean age =  $37.82 \pm 13.3$ ; Gender: 3766 female, 1252 male, 30 other). Currently, there are no ethical guidelines on where global web-based research should be reviewed and whether ethical approval is needed from all countries represented in the study (Looijmans et al., 2022). For studies involving interventions or physical measurements, it is recommended to apply for ethical review in all participating countries, while this is less obvious for online questionnaire studies (Looijmans et al., 2022). Since the current study only involved a brief anonymous online survey and did not include any intervention or physical measurements, we submitted our proposal to The Medical Ethical Review Committee of the Academic Medical Center Amsterdam. The Medical Ethical Review Committee of the Academic Medical Center Amsterdam exempted this study from formal review (W19\_481 # 19.556).

The survey was provided online in 21 languages after consent of participants. The country IP address of participants was used to show the survey in the correct language automatically, but this could also be changed manually by the participants. The survey started with demographic questions about the gender, age, and country of residence of participants. Then, an open text field was presented where respondents were asked to describe the stressful event that currently affected them the most. This was followed by several questions about event characteristics (COVID-19-relatedness, work-relatedness, time since the event took place, whether the event was a repeated/prolonged event and whether it involved physical violence, sexual violence, emotional abuse, serious injury, life-threatening situations, sudden death of a loved one, and/or the respondents causing harm to others). Please note that for the purpose of this paper, only information about work-relatedness (Coded as: 0 = missing; 1 = not work-related event; 2 = work-related event), time since the event took place (Coded as: 0 = missing; 1 = within past month, 2 = one month to half a year ago, 3 = half a year to a year ago, 4 = longer than one year ago), and whether the event was a repeated/prolonged event (Coded as: 0 = missing; 1 = single event; 2 = multiple events /prolonged event) were considered in the present study. Finally, participants filled out the Global Psychotrauma Screen (GPS) questionnaire about the stressful event they previously indicated that currently affects them the most.

### 2.2. Measures

#### 2.2.1. Global Psychotrauma Screen (GPS)

The GPS consists of 22 items about trauma-related symptoms and risk factors with a binary (yes versus no) answer format (Olf et al., 2019). Trauma-related symptoms in the past month are assessed with 17

GPS items and can be summed in a GPS symptom score (range 0-17) with higher scores indicating more severe symptoms. These items include symptoms of PTSD (5 items), disturbances in self-organization (2 items), generalized anxiety disorder (2 items), depressive disorder (2 items), insomnia (1 item), self-harm (1 item), dissociation (2 items), substance abuse (1 item) and other stress-related problems (1 item). A GPS symptom score cut-off of 8 provided optimal sensitivity and specificity for a probable PTSD in previous validation studies (Frewen et al., 2021). The risk factors of the GPS include the occurrence of other stressful events in the past month, lack of social support in the past month, childhood trauma, psychiatric history, and lack of psychological resilience. Previous validation studies indicated that the GPS is a valid and reliable instrument for trauma-related symptoms (Frewen et al., 2021, Oe et al., 2020, Rossi et al., 2021). In the present sample, Cronbach's alpha indicated adequate internal consistency of the GPS symptom score ( $\alpha = .88$ ). Descriptive statistics for the GPS symptom score and risk factor variables are provided in Table S1 in the supplementary material.

### 2.3. Statistical analyses

#### 2.3.1. Language data quality

As noted above, participants taking the survey were asked to provide a written description of the stressful event that currently affected them the most. Participants' written responses were then analyzed to extract information that could predict trauma-related symptoms. For example, we might expect the use of the term abuse to show a positive association with participants' current trauma-related symptoms, regardless of the language used, and thus be useful for prediction purposes. However, the automated methods we used to extract language features can only provide meaningful results when texts are presented in one common language (e.g., the word *abuso*, which is the Italian equivalent of the English word *abuse*, if not translated would be detected and counted as a different word). Because the majority of collected texts were written in English language, we chose the English as a common language framework, thus requiring non-English texts to be translated into English. For all non-English texts, automatic translations were performed using the *googletranslate* function available in the paid-for Google Workspace professional suite linked to an institutional account belonging to one of the authors. Please note that the Google Workspace platform is compliant with EU's General Data Protection Regulation (GDPR), meaning that data confidentiality is enforced on data processed through its translation services. Google makes it clear that "they will not make the content of the text that you send available to the public, or share it with anyone else, except as necessary to provide the Cloud Translation API service"; furthermore, they will not use it for model training purposes (<https://workspace.google.com/learn-more/security/security-whitepaper/page-6.html#limited-data>). Then, authors with the original language as primary language ( $n = 19$ ) checked and rated all translations, corrected the translations when needed, and rated the corrected translations. Translations were rated based on the global evaluation parameters scale (Toledo Báez, 2010). This scale includes a score from 1 (transmission is mostly incoherent) to 5 (transmission is equal to one of an experienced professional translator). The automatic translations by Google Translate were, on average, scored 2.34 ( $SD = .75$ ), which indicated that manual supervision was needed. After corrections, the translations were scored 4.37 ( $SD = .83$ ), which corresponds to satisfactory translations. After these corrections, the text fields included on average 9.21 words ( $SD = 17.21$ ; Range = 1-320; Median = 4).

#### 2.3.2. Feature extraction

In order to extract information that could predict trauma-related symptoms from the text participants provided in the open text field, we extracted two types of features: closed- and open-vocabulary features. Closed-vocabulary analysis was performed on the participants' textual description of stressful events with the English version of the

Linguistic Inquiry Word Count Dictionary (LIWC). For the purpose of this study, we employed the 2007 LIWC dictionary (Pennebaker et al., 2007), consisting of 64 theory-based dictionary categories allowing for the scoring of documents based on the number of words reflecting affective, social, cognitive, and perceptual processes, as well as use of function words such as pronouns, articles, conjunctions, verbs, and adverbs (see for complete list: Pennebaker et al., 2007). For each category, a score is computed as the ratio between the number of words belonging to the specific category over the total number of words in the document.

Open-vocabulary analyses were performed on the textual descriptions of the events and included both topic model analyses and the extraction of n-gram features. First, we implemented a topic model analysis using Latent Dirichlet Allocation (LDA) model. Before running the analyses, we converted all text to lowercase and removed English 'stopwords' (i.e., very frequent words with low specificity), punctuation, and numbers. No stemming was applied to the original text (i.e., words were left unchanged as opposed to being reduced to their word stem, or root form). In order to identify the optimal number of topics, we trained a set of competing LDA models with the following k numbers of topics: 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50. Model training was performed on a random split including 90% of the sample, while 10% were used for model validation. The performance of the competing LDA models was compared by computing the perplexity statistic on the validation set (Wallach et al., 2009); the optimal number of topics was selected using the heuristic approach proposed by Zhao et al. (2015), which is based on examination of the rate of perplexity change (RPC) across LDA models. The coherence of LDA-derived topic-words association was also examined visually using word clouds. Eventually, based on the RPC heuristic procedure and semantic coherence (as assessed using word clouds), we selected  $k = 30$  as the optimal number of topics for this dataset. Word clouds for the extracted topics are shown in the supplementary material (Figs. S1–S30). As the last step, the selected model was applied to all available documents to generate the topic proportion scores. LDA analyses were performed using the Mallet software, version 2.08 (McCallum, 2002).

Extraction of n-grams was performed using the *StringToWordVector* function available in the WEKA software (Hall et al., 2009). Before running the analyses, we converted all text to lowercase and removed numbers, while punctuations were treated as word delimiters. We extracted n-grams ranging from 1 to 3 words that were present in at least 1% of event descriptions (about 51 descriptions). For each event description, counts were obtained for 186 n-grams. For all features extracted from the text description of the stressful event, descriptive statistics are reported in Table S1 in the supplementary material.

#### 2.3.3. Associations of self-report and language features with the GPS symptom score

Pearson's correlation coefficient was used to examine associations between trauma-related symptoms (indicated by the GPS symptom score) and features extracted from the textual description and the self-report section of the event section of the GPS screener. *p*-values were adjusted using a Bonferroni correction (based on total number of examined features) to correct for multiple testing. Please note that categorical predictors (i.e., indicators of gender and event characteristics) were recoded in the form of dummy variables before performing quantitative analyses.

#### 2.3.4. Using Machine learning to predict the GPS symptom score

The relationship between language features and the GPS symptom score was examined separately for each specific subset of language features, namely LIWC, n-grams, topics, and open vocabulary (topics + n-grams), and in a model combining all language features. In order to determine the relative contribution of language features, demographics, event-related characteristics, and risk factors in predicting the GPS symptom score, we tested the models including demographics, event-related characteristics, and risk factors either alone or in combination

with all the language features. First, we examined  $n = 5$  models examining the predictive power of the following language features: (1) LIWC; (2) n-grams; (3) topics; (4) open vocabulary (topics + n-grams), and (5) all language features. Next, for each set of self-report variables (i.e., demographics, event-related characteristics, risk factors, and all self-report information combined), we examined the following models: the model including the specific set of self-report data as features, and the models combining the specific set of self-report data with the five models based on (sub)sets of language features.

The analyses were performed using iterative splitting of the dataset into training and test sets, including respectively 90% and 10% of observations. The training sets were used to train the predictive models, while the test sets were used to evaluate the accuracy of predictions on unseen observations. In order to improve generalizability of results to unseen observations, the splitting procedure and all analyses were repeated 10 times on different train/test splits, and then results were averaged across the 10 splits. Each of the 10 splits was obtained using the *partition()* function of the *splitTools* library in R (Mayer, 2020), which allowed us to generate samples for the train and test sets that were stratified according to the distribution of the GPS symptom score in the overall sample.

Predictive analyses were implemented in the training set using Elastic-Net regression (Zou and Hastie, 2005), Random Forest (Breiman, 2001) regression, and a combination of these models using a stacking ensemble (for a review, see Džeroski and Ženko 2004). The Elastic-Net algorithm is a generalization of linear regression that allows for both parameter regularization and variable selection, thus reducing the risk of overfitting. The Elastic-Net algorithm can be seen as a combination of L1 regularization (i.e., LASSO regression) and L2 regularization (i.e., ridge regression), in which the alpha parameter can be used to weight the contribution of the L1 and L2 regularization, while the lambda parameter controls the amount of parameter shrinkage. For the purpose of this study the alpha parameter was set at 0.001 (i.e., the lowest alpha value allowed in the employed implementation of the Elastic-Net algorithm), while an automated, 10-fold validated search was implemented for the lambda parameter based on an automatically generated sequence of 100 values. By fixing the alpha value to .001, the Elastic-Net model was forced to behave similar to ridge regression. In this way, the algorithm performed both parameter shrinkage and feature selection, but the latter was kept to a minimum. Our choice is related to the need to avoid the major pitfall linked with using large alpha values, which is the risk of underfitting due to the discarding of most features. Additionally, the uncontrolled removal of many features across models would have been problematic because we were interested in comparing the relative importance of different sets of features, or combination of them, for the prediction of the GPS symptom score.

The Random Forest algorithm combines multiple decision trees for the prediction of both continuous and categorical outcomes, and uses averaging to improve the predictive accuracy and control over-fitting. Each decision tree is created using randomly generated samples of the training data. Being based on decision trees, the Random Forest algorithm goes beyond linear regression algorithms by allowing for the automated modelling of potential interactions and non-linearity among features in predicting the outcome. For the purpose of the present study, the Random Forest algorithm was used to perform regression of the GPS continuous score based on 100 trees, with unlimited depth of trees, and based on a random selection of features.

Finally, the stacking ensemble was implemented by combining the Elastic-Net and Random Forest algorithms using a linear regression algorithm with no parameter regularization as a meta-learning algorithm. Using this approach, the meta-learning algorithm combines the predictions from the two machine learning algorithms (i.e., Elastic-Net and Random Forest), possibly improving over the performance of the single algorithm in the ensemble.

The performance of the trained models were evaluated on the test sets. Accuracy of predictions in the test sets was determined with the

correlation between observed and predicted GPS symptom scores and the mean absolute error (MAE) in the predictions, computed as the mean of the absolute differences between observed and predicted scores. The average correlation and MAE values as computed across the test sets are reported as an indication of overall accuracy of the algorithms in predicting the GPS symptom score. Prediction analyses were performed using the Elastic-Net, Random Forest and Stacking algorithm implementation available using the *Rwaka* package (Hornik et al., 2021) for R.

### 2.3.5. Testing the feasibility of a procedure for the classification of clinically relevant symptoms

As a final analytical step, we evaluated the contribution of the extracted features for the purpose of the creation of a screening procedure for detecting clinically relevant trauma-related symptoms with a GPS symptom score of 8 or higher, indicating probable PTSD (Frewen et al., 2021). More specifically, we examined whether the predictive models based on language features alone or in combination with self-report information (i.e., demographics, event-related characteristics, risk factors, and all self-report information combined) could accurately identify those with probable PTSD. Hence, we evaluated the following models: (1) the model including all language features; (2) the model based on self-report demographics and all language features; (3) the model based on event-related characteristics and all language features; (4) the model based on risk factor variables and all language features; (5) the model including all self-report measures and language features. To achieve this aim, in each generated sample fold and for each examined model, we computed the area under the receiver operating characteristic (AUROC) as a measure of association between the predicted GPS (continuous) scores and the binary classification of individuals in PTSD risk groups based on the recording of the observed GPS (continuous) scores using the aforementioned cut-off. To visualize this association, we also generated the ROC curve plots for the folds reporting the best and worst overall classification performance.

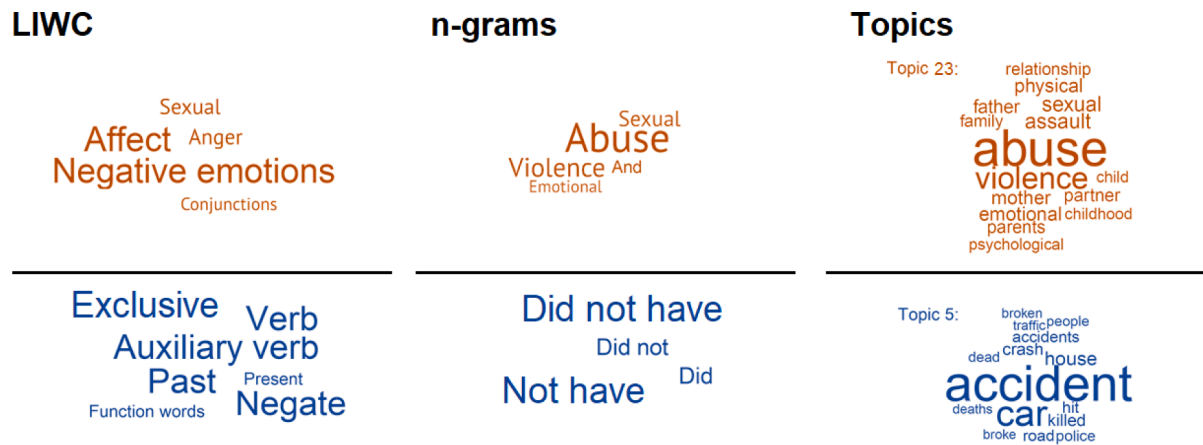
As an additional aim, we evaluated the performance of these models by also comparing their accuracy among groups of participants reporting different numbers of words when describing the event. For this aim, we used a median split (Median = 4 words) to distinguish the groups of participants reporting short descriptions (number of words  $\leq 4$ ) and participants reporting longer descriptions (number of words  $\geq 5$ ).

In accordance with Youngstrom (2014), we considered AUROC values  $\geq .90$  as showing excellent accuracy, values ranging between .80 and .89 as showing good accuracy, values ranging between .70 and .79 as fair accuracy, and values ranging between .60 and .69 as poor accuracy.

## 3. Results

### 3.1. Bivariate associations between self-report and language features with the GPS symptom score

Language features significantly related to GPS symptom scores with a minimal correlation coefficient of  $r = .10$  are shown in Fig. 1 using a word-cloud visualization, while correlation between the GPS symptom scores and all the study measures are reported in Table S2 in the supplementary material. For ease of interpretation, in Fig. 1 features reporting a positive correlation are rendered in red, while those showing negative correlations are rendered in blue. The language feature most strongly positively associated with GPS Symptom scores were topic 23 ( $r = .17, p < .01$ ; top words: *abuse, violence, assault, sexual, emotional, childhood*), followed by the *Negative emotions* LIWC category ( $r = .17, p < .01$ ), the *Affect* LIWC category ( $r = .16, p < .01$ ), the *abuse* n-gram ( $r = .16, p < .01$ ). Other features showing positive correlation with the GPS symptom score were the *Anger* LIWC category ( $r = .13, p < .01$ ), the *Sexual* LIWC category ( $r = .10, p < .01$ ), and the *sexual, and, and violence* n-grams (all showing  $r = .10, p < .01$ ). In turn, the language features



**Note.** Words/n-grams rendered in a bigger font showed stronger positive (red) or negative (blue) associations with the GPS score.

**Fig. 1.** Language features showing strongest positive (red) and negative (blue) correlations with the GPS symptom score

most strongly negatively associated with GPS symptom scores were the *did not have* n-gram ( $r = -.21, p < .01$ ), the *not have* n-gram ( $r = -.20, p < .01$ ), the *Exclusive* ( $r = -.18, p < .01$ ) and *Auxiliary verb* ( $r = -.19, p < .01$ ) LIWC categories, the *Past* LIWC category ( $r = -.18, p < .01$ ), and the *did not* n-gram ( $r = -.17, p < .01$ ). Other features showing negative associations with GPS Symptom score were the *Negate* LIWC category ( $r = -.17, p < .01$ ), the *Verb* LIWC category ( $r = -.16, p < .01$ ), the *did*<sup>2</sup> n-grams ( $r = -.15, p < .01$ ), the *Function words* LIWC category ( $r = -.14, p < .01$ ), the *Present* LIWC category ( $r = -.12, p < .01$ ). Only one topic (i.e., topic 5) showed a significant negative correlation with the GPS Symptom score ( $r = -.10, p < .01$ ; top words: *accident, car, crash, injury, road, people*).

Finally, regarding self-report features, we found that female gender was also associated with higher GPS symptom scores ( $r = .17, p < .01$ ), while being male was associated with lower GPS symptom scores ( $r = -.19, p < .01$ ); in turn, indicating a non-binary gender (i.e., other option) showed no association with the GPS symptom score. Older age was related to lower GPS symptom scores ( $r = -.11, p < .01$ ). Experiencing multiple stressful events was associated with higher GPS symptom scores ( $r = .27, p < .01$ ), while experiencing a single event was associated with lower GPS symptom scores ( $r = -.18, p < .01$ ); in turn, failing to report information about how long ago the event happened was negatively related to the GPS symptom score ( $r = -.23, p < .01$ ). All risk factors were related to higher GPS Symptom scores with small-medium effect sizes (GPS 17 – Stressful events:  $r = .43, p < .01$ ; GPS 19 – Lack of support:  $r = .40, p < .01$ ; GPS 20 – Child trauma:  $r = .27, p < .01$ ; GPS 21 – Psychiatric history:  $r = .33, p < .01$ ; GPS 22 – Resilience:  $r = .10, p < .01$ ).

### 3.2. Machine learning for Prediction of GPS symptom score (number of symptoms)

Next, we assessed the relevance of the language features and self-report information in predicting trauma-related symptoms (as indicated by the GPS symptom score). We first present results of the language features only, followed by the self-report information only and finally the models combining all information.

Results of the tested models are shown in Figs. 2 and 3. For all subsets of features, the stacking ensemble algorithm combining both Elastic-Net and Random Forest algorithms resulted in more accurate predictions compared to models employing these algorithms separately. For this reason, here we only comment the results of models using the stacking

<sup>2</sup> Please note that a Pearson correlation of  $r = .95$  ( $p < .01$ ) was detected in our data between the *did* and *did not* n-grams.

algorithm. Regarding models based on language features, the predictive performance of language was at best when all language features were combined together (average  $r = .37$ ; average MAE = 3.62). Next, sorted in descending order of overall predictive power, were the model based on the open-vocabulary features (average  $r = .35$ ; average MAE = 3.65), the model based including on the n-gram features (average  $r = .35$ ; average MAE = 3.66), and the model including only the LIWC features (average  $r = .34$ , average MAE = 3.67). Finally, the model including only the topic features showed the worst predictive power over the GPS symptom score (average  $r = .27$ ; average MAE = 3.77).

Among models including only self-report data, the model including all self-report variables provided the stronger predictive power (average  $r = .61$ ; average MAE = 2.99), followed by the model including only the risk factors (average  $r = .57$ ; average MAE = 3.14), the models including only event-related characteristics (average  $r = .36$ ; average MAE = 3.63), and demographics (average  $r = .22$ ; average MAE = 3.84) showed the worst predictive power.

Finally, among the models combining language and self-report features, the models including all self-report and language variables showed the best performance (average  $r$  ranging from .62 to .63; average MAE ranging from 2.97 to 2.94), with the model combining all self-report data and n-grams showing the strongest predictive power (average  $r = .63$ ; average MAE = 2.94). All the models combining risk-factors and language variables showed a strong performance (average  $r$  ranging from .58 to .61; average MAE ranging from 3.11 to 3.02), with the model combining risk factors and all language variables showing the highest accuracy (average  $r = .61$ ; average MAE = 3.02). Combining event-related characteristics and language features provided moderate results (average  $r$  ranging from .38 to .42; average MAE ranging from 3.61 to 3.54), with the model combining event-related characteristics and all language variables showing the strongest performance (average  $r = .42$ ; average MAE = 3.54). Finally, models combining demographics and language data showed moderate predictive power (average  $r$  ranging from .31 to .39; average MAE from 3.72 to 3.58), with the model including demographics and all language features showing the best performance (average  $r = .39$ ; average MAE = 3.58).

### 3.3. Testing the feasibility of a procedure for the classification of clinically relevant symptoms

We also assessed whether language features, self-report information and both combined led to accurate predictions of probable PTSD. Table 1 reports the mean, minimum, and maximum AUROC values computed in the randomly generated sample folds representing the association between predicted GPS symptom scores and the binary

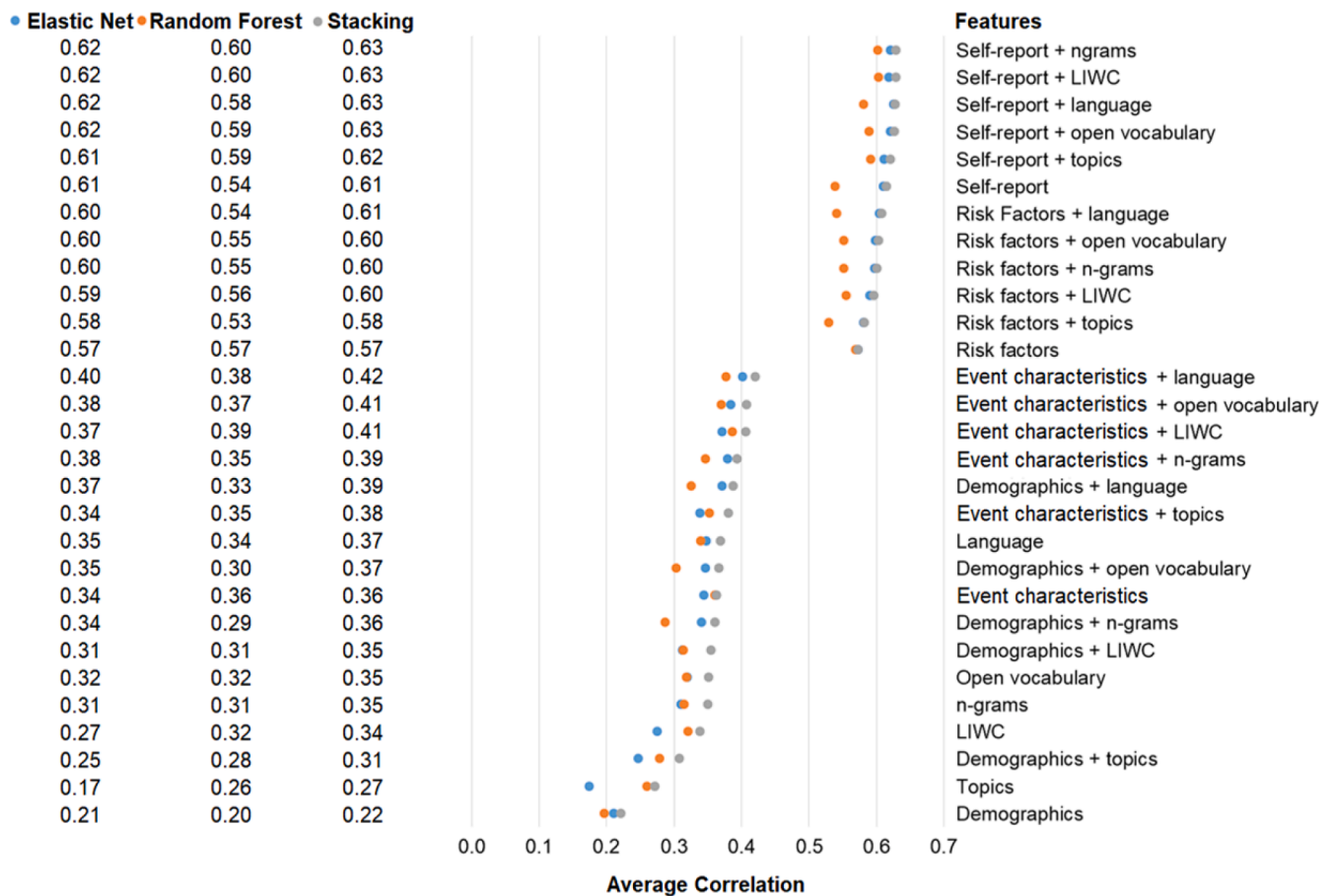


Fig. 2. Average correlation between observed and predicted GPS Symptom score by algorithm and combination of features

classification of individuals in the probable PTSD versus not probable PTSD group. Overall, when examined in the overall sample, none of the models achieved good accuracy in predicting the classification (AUROC  $\geq .80$ ) while the models based on a combination of language and all self-report features achieved fair classification accuracy (average AUROC between .70 and .79). The rest of the models achieved only poor classification accuracy (AUROC between .60 and .69). See Fig. S31 in the supplementary material for visualization of the ROC curves as generated in each fold based on different sets of predictors.

It is worth noting that when split according to the median number of words used by participants in describing the stressful events, we see that performance is slightly improved among participants writing shorter, more concise descriptions (see Table 1), compared to participants providing longer descriptions. Among participants writing descriptions shorter than 5 words, the model combining language features with either risk factors, or all self-report variables, achieved good accuracy in predicting probable PTSD diagnosis (AUROC  $\geq .80$ ).

#### 4. Discussion

In the current study, we explored the feasibility of using short text descriptions provided by participants to describe the stressful event that currently affects them the most for the prediction of trauma-related symptoms and probable PTSD diagnosis. Moreover, we evaluated the feasibility of using a combination of the short text descriptions with a few demographic characteristics and risk factors for predicting trauma-related symptoms. We found that the combination of language features and self-report information resulted in good predictions of trauma-related symptoms and allowed for detection of probable PTSD with a

moderate accuracy. This indicates that short text descriptions are useful for predicting trauma-related symptoms and probable PTSD diagnosis, but only when used alongside some self-report information.

We found that the stacking ensemble algorithm, combining both Random Forest and Elastic-Net regression, resulted in optimal predictive power, improving over the separate application of random forests and Elastic-Net. This is in line with literature indicating that the use of a “super learner” combining multiple algorithms is expected to outperform the use of single algorithms (van der Laan et al., 2007). Furthermore, we found that words, categories, and topics related to (sexual) violence, abuse, and strong (negative) emotions were related to more trauma-related symptoms. In contrast, the topic related to (car) accidents, passive language, and negations were related to less trauma-related symptoms. This is in line with previous literature indicating that interpersonal traumatic events are more strongly related to trauma-related symptoms than non-interpersonal traumatic events (Charuvastra and Cloitre, 2008). A previous study using text mining in patients with and without PTSD also found strong emotions and words referring to interpersonal trauma, such as rape, to be predictive of PTSD (He et al., 2012).

The predictive power of the short text descriptions for trauma-related symptoms was optimal when combining all language features (topics, LIWC, and n-grams). This was comparable to the predictive power of self-reported event characteristics and considerably more than demographic variables. However, since the average correlation between predicted and observed trauma-related symptoms when using only language features was only moderate, this is not sufficient for reliable predictions of trauma-related symptoms at an individual level. In turn, the correlation between observed and predicted trauma-related symptoms was strong in a model using the language features combined with

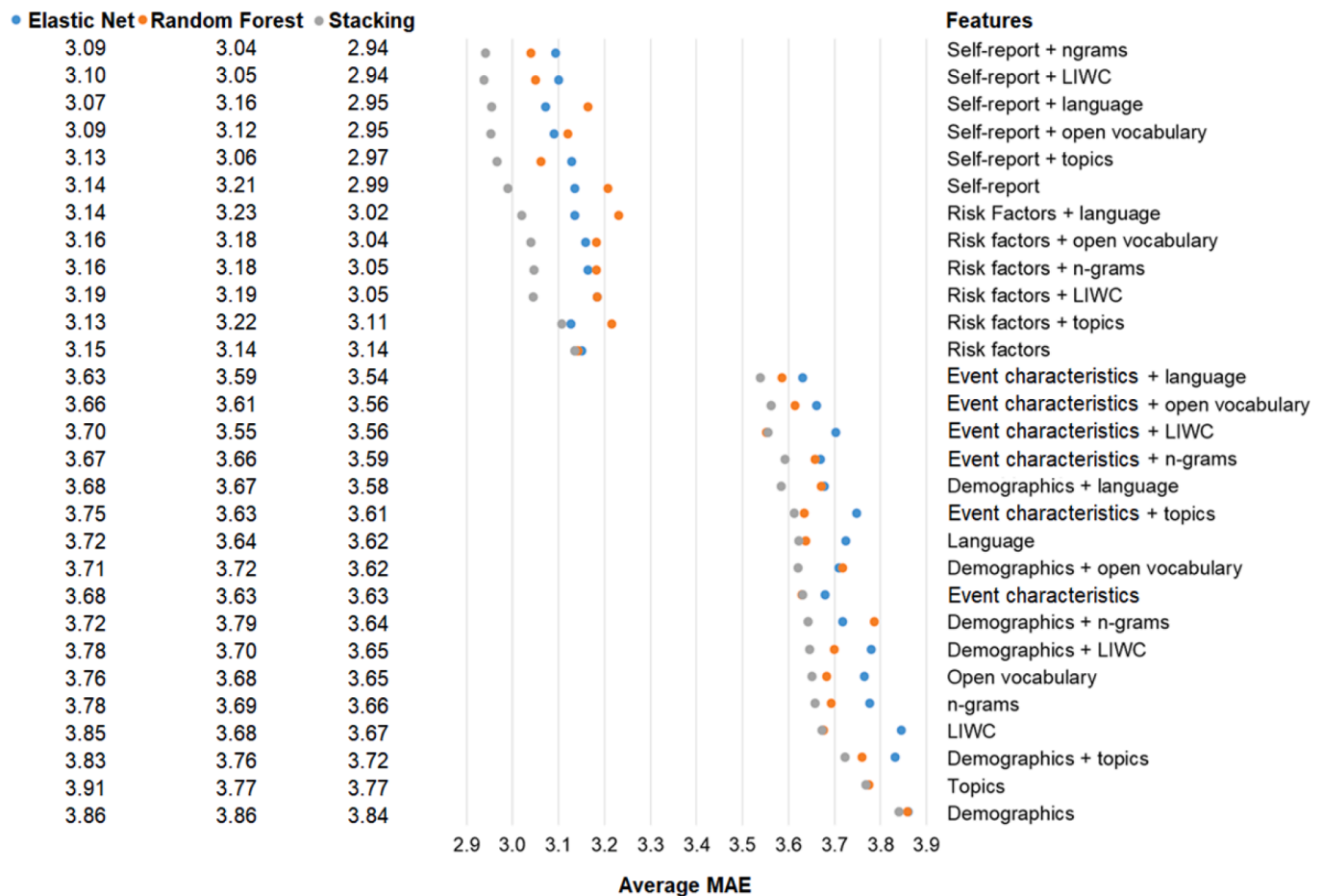


Fig. 3. Average mean absolute error comparing observed and predicted GPS Symptom score by algorithm and combination of features

Table 1

Association between Predicted GPS symptom scores and GPS symptoms classification: Mean, and Range of AUROC values in the overall sample and by length of description

Features	Average AUROC		
	Overall sample Mean (Range)	Description ≤ 4 words Mean (Range)	Description ≥ 5 words Mean (Range)
Language	0.64 (0.61, 0.67)	0.66 (0.59, 0.72)	0.62 (0.59, 0.66)
Demographics + Language	0.66 (0.63, 0.69)	0.67 (0.62, 0.75)	0.64 (0.60, 0.71)
Event characteristics + Language	0.67 (0.64, 0.70)	0.69 (0.64, 0.74)	0.66 (0.61, 0.71)
Risk factors + Language	0.78 (0.76, 0.81)	0.80 (0.75, 0.83)	0.76 (0.73, 0.79)
All Self-Report + Language	0.79 (0.76, 0.82)	0.81 (0.76, 0.83)	0.77 (0.74, 0.80)

demographic information, event-related characteristics, and risk factors, pointing toward a possible use for screening purposes.

We also tested the performance of the aforementioned models in detecting probable PTSD. Based on AUROC values, only the model including language features combined with demographic information, event-related characteristics, and risk factors resulted in detection of probable PTSD with a moderate accuracy (accuracy = .79). Notably, it has been argued that in the field of mental health, reaching an AUROC value of 1 might be impossible because the criterion diagnosis is imperfect, thereby constraining AUROC values. Therefore, an AUROC

value between .70 and .80 might actually represent a good accuracy (Youngstrom, 2014). Interestingly, a previous study using textual information found an even higher accuracy in detecting PTSD (accuracy = .84; He et al., 2019). In this study, participants were recommended to write at least 150 words about their traumatic event and their symptoms (He et al., 2019). Hence, the language information from this study was likely to include much more information. In the present study, we found evidence of a stronger accuracy among participants reporting very short descriptions (number of words ≤ 4 words), when compared with participants providing longer descriptions. Among participants reporting very short descriptions, the model combining language features with either all self-report features, or just risk factors, showed improved accuracy (accuracy ≥ .80). This finding highlights the potential of the automated coding of free response texts, including those with very short answers, for the purpose of the detection of symptoms of distress related to stressful events. Indeed, the ability of respondents to express freely about their personal experience may help in extending the spectrum of difficulties assessed by self-report questionnaire based on predetermined rating scale questions, ultimately improving the overall validity of assessment (Kjell et al., 2019). One limitation related to use of open-ended questions lays in their coding procedure, which typically requires additional resources compared to that of closed-ended questions; however, recent advances in the use of automated coding approaches, such as those used in the present study, point toward the possibility of a wider use of this type of open-ended questions (Kjell et al., 2019). Future studies might try to find the optimal trade-off between the length of the text response and the burden of time and emotional discomfort for participants.

The results of the current study have several implications for efforts

to improve accurate screening for trauma-related disorders. We found that for accurate predictions, the open text field and only 10 items were needed, including very general information such as gender, whether the event was single or repeated or whether someone experienced social support. Thus, the burden for participants to fill out these items was minimal, especially compared to reporting a range of mental health symptoms or a clinical interview. The open text field did not require much from participants as we did not instruct participants to write a large number of words, nor did we limit the number of words. In fact, the way the participants decided to fill out the question (number of words, active or passive form, emotional or not emotional etc.) provided relevant information for the prediction of their trauma-related symptoms. This makes the screening process very efficient without posing any burden or requiring a lot of effort from participants. The combination of some self-report information and textual information might be used to construct an online screening instrument which provides personal feedback for people about their trauma-related symptoms and the likelihood that they meet criteria for a trauma-related disorder. The current results suggest that this is feasible for PTSD. It might be worthwhile to investigate whether we can further improve the accuracy to avoid false positives and false negatives by adding a few more open-ended questions. For example, one might ask participants to describe the symptoms that affect them most. It might also be interesting to add a question about treatment expectancies and preferences, especially since this might also be relevant for personalized treatment indications. Future studies might investigate whether such questions can be used to also predict other trauma-related disorders, as open-text fields might be less disorder-specific compared to standard screening questionnaires. This might further optimize efficient screening for trauma-related disorders.

In the current study, we used data from 21 languages, and we tried to translate this data into English automatically. When native speakers systematically rated this data, the automatic translations were rated mediocre and had to be manually supervised. This indicates that it is not possible to rely on automatic translations using Google Translate at the moment. Future studies might investigate other automated translation options as this would greatly reduce the amount of manual labor. Alternatively, it would be interesting to confirm our findings using non-English language texts.

This study has several limitations. Firstly, we did not establish PTSD diagnoses using a clinician-administered interview. This would have allowed for more reliable information about the potential of the models to detect PTSD diagnoses. Secondly, we included a limited set of predictors in this study. Although this resulted in good predictions of trauma-related symptoms, we do not know whether and how many additional predictors (e.g. peritraumatic distress; Vance et al., 2018) would have increased the performance. Thirdly, although the large, diverse global sample is a strength of the current study and increases the generalizability to diverse countries and languages, this sample was recruited via the internet and, therefore, a convenience sample. Finally, the translations of the texts into English might have slightly changed the context and meaning of sentences and changed the number of words.

Despite these limitations, this is the first study that evaluated whether short text descriptions about the traumatic event might be useful in predicting trauma-related symptoms and probable PTSD in a large cross-national sample. We found that combining information from these text descriptions with demographic information, event-related characteristics, and risk factors led to good predictions of trauma-related symptoms and moderate prediction of probable PTSD diagnosis. We conclude that further research into the use of short text descriptions about the traumatic event is worthwhile, especially in combination with other predictors.

#### CRedit authorship contribution statement

**D. Marengo:** Conceptualization, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft. **C.M.**

**Hoeboer:** Conceptualization, Methodology, Project administration, Writing – original draft. **B.P. Veldkamp:** Conceptualization, Methodology, Supervision, Writing – review & editing. **M. Olff:** Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing.

#### Declaration of Competing Interest

There are no conflicts of interest to declare that are relevant to the content of this article.

#### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

#### Data statement

Raw data were generated through the Global Collaboration on Traumatic Stress (GC-TS). Derived data supporting the findings of this study are available on the GC-TS website: [www.global-psychotrauma.net](http://www.global-psychotrauma.net).

#### Acknowledgment

We thank all ambassadors who helped include participants around the world, in particular: Helene Aakvaag, Dean Ajdukovic, Zafer Altunbezel, Xenia Anastassiou-Hadjicharalambous, Vittoria Ardino, Anne Bakker, Sara Belquaid, Jonathon Bisson, Erine Bröcker, Lucia Cantoni, Ruby Charak, Marylene Cloitre, Aida Dias, Malgorzata Dragan, Atle Dyregrov, Julian Ford, Sarah Gallo, Paul Frewen, Wissam El Hage, Juanita Haagsma, Jackie June ter Heide, Danielle Hett, Maryke Hewett, Jana Javakhishvili, Nancy Kassam-Adams, Yoshiharu Kim, Christian Kristensen, Rachel Langevin, Juliana Lanza, Patrick Lorenz, Weili Lu, Brigitte Lueger-Schuster, Sam Manickam, Marcelo Mello, Gladys Mwiti, Natalia Nalyvaiko, Angela Nickerson, Misari Oe, Heval Özgen, Janaina Pinto, Daniela Rabellino, Luisa Sales, Carolina Salgado, Julia Schellong, Ulrich Schnyder, Soraya Seedat, Nadejda Semenova, Andrew Smith, Sjacko Sobczak, Erik de Soir, Zhonglin Tan, Keerthana Thatavarthi, Carmelo Vazquez, Anne Wagner, Li Wang, Irina Zrnica.

#### Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.psychres.2022.114753](https://doi.org/10.1016/j.psychres.2022.114753).

#### References

- APA, 2013. Diagnostic and statistical manual of mental disorders (5th ed.). American Psychiatric Association.
- Arora, T., Grey, I., Ostlundh, L., Lam, K.B.H., Omar, O.M., Arnone, D., 2020. The prevalence of psychological consequences of COVID-19: a systematic review and meta-analysis of observational studies. *J. Health Psychol.* <https://doi.org/10.1177/1359105320966639>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Breslau, N., Davis, G.C., Peterson, E.L., Schultz, L.R., 2000. A second look at comorbidity in victims of trauma: the posttraumatic stress disorder-major depression connection. *Biol. Psychiatry* 48 (9), 902–909. [https://doi.org/10.1016/S0006-3223\(00\)00933-1](https://doi.org/10.1016/S0006-3223(00)00933-1).
- Charuvastra, A., Cloitre, M., 2008. Social bonds and posttraumatic stress disorder. *Annu. Rev. Psychol.* 59, 301–328. <https://doi.org/10.1146/annurev.psych.58.110405.085650>.
- Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., Dobson, K.S., 2013. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *Can. J. Psychiatry Revue Canadienne De Psychiatrie* 58 (7), 376–385. <https://doi.org/10.1177/070674371305800702>.
- de Vries, G.J., Olff, M., 2009. The lifetime prevalence of traumatic events and posttraumatic stress disorder in the Netherlands. *J. Trauma. Stress* 22 (4), 259–267. <https://doi.org/10.1002/jts.20429>.



- Denckla, C.A., Gelaye, B., Orlinsky, L., Koenen, K.C., 2020. REACH for mental health in the COVID19 pandemic: an urgent call for public health action. *Eur. J. Psychotraumatol.* 11 (1) <https://doi.org/10.1080/20008198.2020.1762995>.
- Džeroski, S., Ženko, B., 2004. Is combining classifiers with stacking better than selecting the best one? *Mach. Learn.* 54 (3), 255–273. <https://doi.org/10.1023/B:MACH.0000015881.36452.6e>.
- Frewen, P., McPhail, I., Schnyder, U., Oe, M., Olf, M., 2021. Global psychotrauma screen (gps): psychometric properties in two internet-based studies. *Eur. J. Psychotraumatol.* 12 (1) <https://doi.org/10.1080/20008198.2021.1881725>.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. *SIGKDD Explor. News.* 11 (1), 10–18. <https://doi.org/10.1145/1656274.1656278>.
- He, Q.W., Veldkamp, B.P., de Vries, T., 2012. Screening for posttraumatic stress disorder using verbal features in self narratives: a text mining approach. *Psychiatry Res.* 198 (3), 441–447. <https://doi.org/10.1016/j.psychres.2012.01.032>.
- He, Q.W., Veldkamp, B.P., Glas, C.A.W., van den Berg, S.M., 2019. Combining text mining of long constructed responses and item-based measures: a hybrid test design to screen for posttraumatic stress disorder (PTSD). *Front. Psychol.* 10, 2358, ARTN. <https://doi.org/10.3389/fpsyg.2019.01811>.
- Hornik, K., Buchta, C., Hothorn, T., Karatzoglou, A., Meyer, D., and Zeileis, A. (2021). *Package 'RWeka'*. Retrieved 25-01-2022 from <https://cran.r-project.org/web/packages/RWeka/RWeka.pdf>.
- Kessler, R.C., Aguilar-Gaxiola, S., Alonso, J., Benjet, C., Bromet, E.J., Cardoso, G., Degenhardt, L., de Girolamo, G., Dinolova, R.V., Ferry, F., Florescu, S., Gureje, O., Haro, J.M., Huang, Y.Q., Karam, E.G., Kawakami, N., Lee, S., Lepine, J.P., Levinson, D., Navarro-Mateu, F., Pennell, B.E., Piazza, M., Posada-Villa, J., Scott, K.M., Stein, D.J., Ten Have, M., Torres, Y., Viana, M.C., Petukhova, M.V., Sampson, N.A., Zaslavsky, A.M., Koenen, K.C., Survey, W.W.M.H., 2017. Trauma and PTSD in the WHO world mental health surveys. *Eur. J. Psychotraumatol.* 8 <https://doi.org/10.1080/20008198.2017.1353383>.
- Kessler, R.C., Green, J.G., Gruber, M.J., Sampson, N.A., Bromet, E., Cuitan, M., Furukawa, T.A., Gureje, O., Hinkov, H., Hu, C.Y., Lara, C., Lee, S., Mneimneh, Z., Myer, L., Oakley-Browne, M., Posada-Villa, J., Sagar, R., Viana, M.C., Zaslavsky, A.M., 2010. Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. *Int. J. Methods Psychiatr. Res.* 19, 4–22. <https://doi.org/10.1002/mpr.310>.
- Kjell, O.N., Kjell, K., Garcia, D., Sikström, S., 2019. Semantic measures: using natural language processing to measure, differentiate, and describe psychological constructs. *Psychol. Methods* 24 (1), 92. <https://doi.org/10.1037/met0000191>.
- Kuhar, M., Kocjan, G.Z., 2021. Associations of adverse and positive childhood experiences with adult physical and mental health and risk behaviours in Slovenia. *Eur. J. Psychotraumatol.* 12 (1). ArtN1924953 [10.1080/20008198.2021.1924953](https://doi.org/10.1080/20008198.2021.1924953).
- Looijmans, A., Spahrkas, S.S., Sanderman, R., Hagedoorn, M., 2022. Ethical review procedures in international internet-based intervention studies. *Internet Interv. Appl. Inf. Technol. Ment. Behav. Health* 28. ARTN100487. <https://doi.org/10.1016/j.neubiorev.2021.03.024>.
- Magruder, K.M., McLaughlin, K.A., Borbon, D.L.E., 2017. Trauma is a public health issue. *Eur. J. Psychotraumatol.* 8 (1). ArtN 1375338 [10.1080/20008198.2017.1375338](https://doi.org/10.1080/20008198.2017.1375338).
- Marvaldi, M., Mallet, J., Dubertret, C., Moro, M.R., Guessoum, S.B., 2021. Anxiety, depression, trauma-related, and sleep disorders among healthcare workers during the COVID-19 pandemic: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 126, 252–264. <https://doi.org/10.1016/j.neubiorev.2021.03.024>.
- Mavranzouli, I., Megnin-Viggars, O., Daly, C., Dias, S., Welton, N.J., Stockton, S., Bhutani, G., Grey, N., Leach, J., Greenberg, N., Katona, C., El-Leithy, S., Pilling, S., 2020. Psychological treatments for post-traumatic stress disorder in adults: a network meta-analysis. *Psychol. Med.* 50 (4), 542–555. <https://doi.org/10.1017/S0033291720000070>.
- Mayer, M. (2020). *Package 'splitTools'*. Retrieved 25-01-2022 from <https://cran.r-project.org/web/packages/splitTools/splitTools.pdf>.
- McCallum, A. K. (2002). *MALLET: a machine learning for language toolkit*. <http://mallet.cs.umass.edu/>.
- Milanak, M.E., Zuromski, K.L., Cero, I., Wilkerson, A.K., Resnick, H.S., Kilpatrick, D.G., 2019. Traumatic event exposure, posttraumatic stress disorder, and sleep disturbances in a national sample of U.S. adults. *J. Trauma. Stress* 32 (1), 14–22. <https://doi.org/10.1002/jts.22360>.
- Oe, M., Kobayashi, Y., Ishida, T., Chiba, H., Matsuoka, M., Kakuma, T., Frewen, P., Olf, M., 2020. Screening for psychotrauma related symptoms: Japanese translation and pilot testing of the global psychotrauma screen. *Eur. J. Psychotraumatol.* 11 (1) <https://doi.org/10.1080/20008198.2020.1810893>.
- Olf, M., Amstadter, A., Armour, C., Birkeland, M.S., Bui, E., Cloitre, M., Ehlers, A., Ford, J.D., Greene, T., Hansen, M., Lanius, R., Roberts, N., Rosner, R., Thoresen, S., 2019. A decennial review of psychotraumatology: what did we learn and where are we going? *Eur. J. Psychotraumatol.* 10 (1) <https://doi.org/10.1080/20008198.2019.1672948>.
- Olf, M., Bakker, A., Frewen, P., Aakvaag, H., Adjukovic, D., Brewer, D., Borbon, D.L.E., Cloitre, M., Hyland, P., Kassam-Adams, N., Knefel, M., Lanza, J.A., Lueger-Schuster, B., Nickerson, A., Oe, M., Pfaltz, M.C., Salgado, C., Seedat, S., Wagner, A., Schnyder, U., Bisson, J.I., Dyb, G.A., Hasan, A., Hagh, S.A.M., Mwit, G.K., Pinto, J., Salimi, Y., Schafer, I., Schellong, J., Shigemura, J., Anastassiou-Hadjicharalambous, X., Primasari, I., Ge-Ts, 2020. Screening for consequences of trauma - an update on the global collaboration on traumatic stress. *Eur. J. Psychotraumatol.* 11 (1) <https://doi.org/10.1080/20008198.2020.1752504>.
- Olf, M., Primasari, I., Qing, Y.L., Coimbra, B.M., Hovnanyan, A., Grace, E., Williamson, R.E., Hoebner, C.M., Consortium, G.-C., 2021. Mental health responses to COVID-19 around the world. *Eur. J. Psychotraumatol.* 12 (1). ArtN 1929754. <https://doi.org/10.1080/20008198.2021.1929754>.
- Pagotto, L.F., Mendlowicz, M.V., Coutinho, E.S.F., Figueira, I., Luz, M.P., Araujo, A.X., Berger, W., 2015. The impact of posttraumatic symptoms and comorbid mental disorders on the health-related quality of life in treatment-seeking PTSD patients. *Compr. Psychiatry* 58, 68–73. <https://doi.org/10.1016/j.comppsy.2015.01.002>.
- Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A., Booth, R.J., 2007. *The Development and Psychometric Properties of LIWC*. Lawrence Erlbaum Associates.
- Ramos-Lima, L.F., Waikamp, V., Antonelli-Salgado, T., Passos, I.C., Freitas, L.H.M., 2020. The use of machine learning techniques in trauma-related disorders: a systematic review. *J. Psychiatr. Res.* 121, 159–172. <https://doi.org/10.1016/j.jpsychires.2019.12.001>.
- Rossi, R., Soccì, V., Talevi, D., Niolu, C., Pacitti, F., Di Marco, A., Rossi, A., Siracusano, A., Di Lorenzo, G., Olf, M., 2021. Trauma-spectrum symptoms among the Italian general population in the time of the COVID-19 outbreak. *Eur. J. Psychotraumatol.* 12 (1) <https://doi.org/10.1080/20008198.2020.1855888>.
- Toledo Báez, M.C., 2010. *El resumen automático y la evaluación de traducciones en el contexto de la traducción especializada*. Peter Lang International Academic Publishers.
- Trauer, J.M., Qian, M.Y., Doyle, J.S., Rajaratnam, S.M.W., Cunningham, D., 2015. Cognitive behavioral therapy for chronic insomnia: a systematic review and meta-analysis. *Ann. Intern. Med.* 163 (3) <https://doi.org/10.7326/M14-2841>, 191–+.
- van der Laan, M.J., Polley, E.C., Hubbard, A.E., 2007. Super learner. *Stat. Appl. Genet. Mol. Biol.* 6.
- Vance, M.C., Kovachy, B., Dong, M., Bui, E., 2018. Peritraumatic distress: a review and synthesis of 15 years of research. *J. Clin. Psychol.* 74 (9), 1457–1484. [10.1002/jclp.22612](https://doi.org/10.1002/jclp.22612).
- Wallach, H.M., Murray, I., Salakhutdinov, R., Mimno, D., 2009. In: *Evaluation methods for topic models*. Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Quebec, Canada. <https://doi.org/10.1145/1553374.1553515>.
- Youngstrom, E.A., 2014. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *J. Pediatr. Psychol.* 39 (2), 204–221. <https://doi.org/10.1093/jpepsy/jst062>.
- Zhao, W., Chen, J.J., Perkins, R., Liu, Z., Ge, W., Ding, Y., Zou, W., 2015. A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinf.* 16 (13), S8. <https://doi.org/10.1186/1471-2105-16-S13-S8>.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* 67 (2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.