Utilization of Machine Learning-Based Computer Vision and Voice Analysis to Derive Digital Biomarkers of Cognitive Functioning in Trauma Survivors

Katharina Schultebraucks a, b, c  Vijay Yadav d  Isaac R. Galatzer-Levy b, d

a Vagelos School of Physicians and Surgeons, Department of Emergency Medicine, Columbia University Medical Center, New York, NY, USA; b Department of Psychiatry, New York University School of Medicine, New York, NY, USA; c Data Science Institute, Columbia University, New York, NY, USA; d AiCure, New York, NY, USA

Keywords
Computer vision · Deep learning · Voice analysis · Emergency department · Digital biomarkers · Cognitive functioning

Abstract
Background: Alterations in multiple domains of cognition have been observed in individuals who have experienced a traumatic stressor. These domains may provide important insights in identifying underlying neurobiological dysfunction driving an individual’s clinical response to trauma. However, such assessments are burdensome, costly, and time-consuming. To overcome barriers, efforts have emerged to measure multiple domains of cognitive functioning through the application of machine learning (ML) models to passive data sources. Methods: We utilized automated computer vision and voice analysis methods to extract facial, movement, and speech characteristics from semi-structured clinical interviews in 81 trauma survivors who additionally completed a cognitive assessment battery. A ML-based regression framework was used to identify variance in visual and auditory measures that relate to multiple cognitive domains. Results: Models derived from visual and auditory measures collectively accounted for a large variance in multiple domains of cognitive functioning, including motor coordination ($R^2 = 0.52$), processing speed ($R^2 = 0.42$), emotional bias ($R^2 = 0.52$), sustained attention ($R^2 = 0.51$), controlled attention ($R^2 = 0.44$), cognitive flexibility ($R^2 = 0.43$), cognitive inhibition ($R^2 = 0.64$), and executive functioning ($R^2 = 0.63$), consistent with the high test-retest reliability of traditional cognitive assessments. Face, voice, speech content, and movement have all significantly contributed to explaining the variance in predicting functioning in all cognitive domains. Conclusions: The results demonstrate the feasibility of automated measurement of reliable proxies of cognitive functioning through low-burden passive patient evaluations. This makes it easier to monitor cognitive functions and to intervene earlier and at a lower threshold without requiring a time-consuming neurocognitive assessment by, for instance, a licensed psychologist with specialized training in neuropsychology.

Introduction

Accurate measurement of cognitive performance is key for selecting treatment options, the tracking of treatment response, and identifying risk factors for the development of adverse effects across diverse neuropsychiatric disorders developing after trauma exposure [1–8].
Digital Biomarkers of Cognitive Functioning

Digital Biomarkers of Cognitive Functioning has been linked to posttraumatic stress disorder (PTSD) symptoms in trauma survivors [1, 2] and may thereby impede the processing of traumatic memories. Impaired cognitive functioning after trauma is also associated with other trauma-related psychiatric disorders such as depression [4] and PTSD [1, 2, 6, 7]. Cognitive impairment can also be a risk factor for later developing psychopathologies [3] as well as an early sign of already developed psychopathologies. However, standard assessment methods for cognitive evaluation are burdensome to subjects and clinicians, cost-prohibitive, and allow for only infrequent assessment due to potential practice effects. Barriers include access to professionals to conduct testing and the cost of testing itself, along with the time and effort of the evaluation. While computerized batteries have addressed some of the prohibitive financial aspects, they still require significant time and effort. Further, they cannot be administered at the frequency needed to assess short-term fluctuations in cognitive status. Increasingly there is a need for frequent, noninvasive, and low-burden methods to measure cognitive functioning to support research and clinical decision-making in diverse neuropsychiatric populations.

Multiple methodologies for passive measurement of cognitive functioning have emerged that utilize digital data sources [9, 10]. These methods utilize digital information such as cell phone sensor data or keystroke activity, alongside traditional cognitive tests, to develop machine learning (ML)-based algorithms [9, 11]. These algorithms identify shared variance between the traditional metrics and the passive data source, resulting in a high-frequency passive metric that is informative about cognitive functioning [11]. These digital phenotyping methodologies hold promise to significantly lower the assessment burden and increase the frequency and sensitivity of cognitive measurement in clinical populations. Consistent with other technology-based solutions to previously analog evaluation methods such as web-based neurocognitive assessment batteries [12], such approaches hold promise to greatly increase the scale and reduce the costs associated with cognitive evaluation, ultimately advancing treatment development, risk assessment, allocation of treatment resources, and outcome assessment. Digital phenotyping has been validated in numerous settings as a secure and safe method of assessment [13] and can be used for affect detection with high accuracy [14].

Digital phenotyping integrates behavioral information from multiple modalities using passively observed facial, movement, speech, and voice data [10]. A significant body of research has identified facial, vocal, and motor movement markers of neuropsychiatric functioning [15–17]. Theoretically informed data-driven computational digital phenotyping constitutes an innovative approach to examining symptom dimensions in line with the National Institute of Mental Health research domain criteria [10].

Methodologies in computer vision and voice acoustic analysis increasingly allow for the automated measurement of facial affect and its intensity, movement, and speech acoustics. These domains allow for the quantification of features that reflect domains such as affect, rate of speech, patterns of eye contact, and motor activity that may serve as informative proxies of multiple domains of cognitive functioning. In the current work, we determined whether multiple domains of cognitive and motor functioning can be accurately measured using audio and video data sources to collectively predict functioning in these key domains of cognition.

Methods

Participants

Data were taken from a prospective longitudinal cohort study that enrolled trauma survivors of a DSM-5 PTSD criterion A trauma who were admitted to the emergency department (ED) of a level-1 trauma center (n = 221). Participants were enrolled when admitted to the ED of the Bellevue Hospital Center, New York City from 2012 to 2017. For further inclusion and exclusion criteria please see Schultebraucks et al. [18].

Among those recruited, 36.65% (n = 81) participated in the administration of an automated cognitive battery along with a brief qualitative interview about their experience. Approximately 55.6% of those experienced a head trauma resulting in CT evaluation during emergency care. Participants demonstrated a broad range of clinical symptom severity and cognitive functioning (Table 1).

Procedure

Participants who were recruited in the ED were invited to complete laboratory-based procedures 1 month after ED discharge.

Qualitative Interview. Participants were led through a brief qualitative interview while being audio and video recorded. Video cameras captured participants face-on, with their head and shoulders in the frame. Patients were asked to respond to five questions within a predetermined time limit of 3 min for each of the following questions: (1) "Tell me about your life before the event that brought you to the hospital." (2) "Tell me about the event that brought you to the hospital." (3) "Tell me about your hospital experience." (4) "Tell me about your life since leaving the hospital." (5) "What are your expectations about life in the future?" Interviewers only used brief predetermined follow-up questions when patients stopped responding before 3 min (i.e., such as "could you tell me more about that?").
Table 1. Demographic statistics of all participants (n = 81)

| Demographic Variable                          | Mean ± SD       |
|----------------------------------------------|-----------------|
| Age, years                                   | 37.86 ± 13.99   |
| Female sex, %                                | 42.5            |
| Education, years (mean ± SD)                 | 14.75 ± 3.39    |
| PCL-5 (mean ± SD)                            | 27.30 ± 16.42   |
| Provisional PTSD diagnosis (%)               | 31.6            |
| CES-D (mean ± SD)                            | 21.06 ± 11.73   |
| Provisional depression diagnosis (%)         | 41.8            |
| Loss of consciousness (%)                   | 29.6            |
| Administration of head CT (%)               | 55.6            |

A PCL-5 total score ≥33 and a CES-D cutoff ≥23 was defined as the cutoff for screening positive for a provisional diagnosis of PTSD [21] and a provisional depression diagnosis [40]. CES-D, Center for Epidemiological Studies of Depression Scale; PCL-5, Posttraumatic Stress Disorder Checklist for DSM-5; PTSD, post-traumatic stress disorder.

Cognitive Evaluation. We performed the WebNeuro™ [12] test battery 1 month after ED discharge to measure cognitive functioning. WebNeuro™ is a widely used and validated web-based cognitive assessment battery that is validated against clinician-administered, in-person gold standard cognitive assessments [12, 19]. WebNeuro™ is the web-based version of the IntegNeuro computerized battery that was developed based on the standardized Brain Resource International Database (BRID). The BRID aimed to create a standardized testing approach and integrated them with other sources of data such as genetic, neuroimaging, and clinical data. The strength of WebNeuro™ is that its data are linked to the BRID, which allows to get standardized, highly valid neurocognitive functioning norms [19]. WebNeuro™ has been used in other recent studies, for instance to examine cognitive flexibility [7] or to investigate neurocognitive functioning as vulnerability factors for the development of PTSD [3, 20]. The test battery contains digit span (working memory), a memory recognition task (recall memory), a Stroop task (response inhibition), a maze task (executive function), an emotional identification task, an emotional bias task, a go/no-go task (response inhibition), a motor tapping task (motor coordination), as well as a switching of attention task (cognitive flexibility). Cognitive functioning was calculated automatically by WebNeuro™ and provided a standardized Z score for each participant for the eight neurocognitive composite domains: motor coordination, processing speed, sustained attention, controlled attention, cognitive flexibility, response inhibition, executive function, and emotional bias.

Clinical Measures. Patient-reported clinical assessments of PTSD using the PTSD Checklist for DSM-5 [21] and of depression measured with the Center for Epidemiological Studies of Depression Scale [22] were administered. Further, evidence of head injury and administration of CT scans was evaluated based on the Bellevue Hospital's electronic medical record system Epic.

Statistical Analysis: Extraction of Digital Biomarkers

Video Processing. We used OpenFace [23], a Python library, to process raw videos where each frame was extracted and then broken down to 3×m×n matrices, with m rows and n columns representing red, green, and blue spectrum information. Each value in each m×n matrix represents pixel values from light to dark on the corresponding color spectrum to get raw-level frame-by-frame facial features. Facial features refer to the variables that encode changes in color and brightness reflecting the motion and activation of muscles under the skin of a face during the recorded qualitative interview. OpenFace outputs the action unit intensity and action unit presence value which encodes facial muscle activation. It also computes facial artifacts corresponding to action units based on the Facial Action Coding System (FACS) encoding [24]. Based on these facial artifacts, a facial expressivity score is computed for each emotion (happiness, sadness, anger, disgust, surprise, fear, contempt), a prevalence score for each emotion, and peak expressivity (1, 3, 6, 9, 12, 15 s) windows. We calculated the normalized Emotional FACS (EMFACS) emotions (per frame) to label emotions in the videos. In addition, we calculated the facial expressivity index, which is a composite score of general expressivity that is derived from a subset of action units. For calculating the facial expressivity index, all normalized EMFACS emotion are averaged to determine whether a participant elicits any emotion (independent of which emotion). Furthermore, the expressivity peak count was calculated by adding all peaks across the seven EMFACS-coded emotions for a specified time window.

Acoustic Processing. We used Parselmouth [25], a Python library (implementation of the Praat software [26]), to process raw audio files where each frame was extracted based on the temporal resolution of audio signals. We calculated the audio expressivity index by adding up intensity, formant, pitch variability, and normalized amplitude quotient. In addition, we calculated audio intensity (dB), fundamental frequency (Hz; lowest frequency of the periodic waveform), the harmonic noise ratio (i.e., degree of acoustic periodicity of audio wave/the ratio between periodic and aperiodic components), as well as the glottal to noise excitation ratio (ratio between the noise generated by vocal fold and vocal tract). Furthermore, we calculated the voice frame score (i.e., the audio frame where participants are speaking considering the entire audio segment) and the formant frequency variability which is clustering acoustic energy at a particular frequency in the audio stream. Moreover, we calculated intensity variability (i.e., the variability of audio intensity) and pitch variability (i.e., variation in fundamental frequency frame by frame) as well as the normalized amplitude quotient (i.e., the ratio between ac flow amplitude and negative peak amplitude of ac flow which is normalized w.r.t. to period).

Content Processing. We used Deep Speech [27], a pre-trained deep learning model, to convert audio to text. It extracts features such as rate of speech, intent expressivity, emotion label, and word repetition.

Movement Processing. Movement features were extracted from raw MP4 video files utilizing the OpenFace package [13]. We calculated head movement (i.e., overall body motion), attentiveness (i.e., attention of the participant towards camera within the video frame; this is analyzed by investigating the eye gaze direction coordinates averaged for both eyes), as well as the pupil dilation rate (i.e., change in pupil size within the video frames). Furthermore, we examined head rotation, i.e., rotation within the scene is calculated by observing the rotation (in radians) around the x, y, and z axis. This can be interpreted as a combination of pitch, yaw, and roll with the camera being the origin. It is the displacement in head position by calculating the Euclidean distance of pitch, yaw, and
roll frame by frame (online suppl. Fig. 1; for all online suppl. material, see www.karger.com/doi/10.1159/000512394).

For more information on the extraction of the video, audio, content, and movement features, please see Schultebraucks et al. [17]. We successfully processed all available video and audio data from all participants. No data were lost during processing.

Model Development and Evaluation

For data preprocessing, we normalized categorical variables by using one-hot-encoding and normalization of numerical variables to the value range between 0 and 1. We removed variables with near-zero variance and imputed the data for missing values (±1% values) using median imputation [28].

To reduce the risk of overtraining, we split the dataset into a training and a test sample. The test sample was defined as a hold-out sample which was not part of the training. We then used the feature selection algorithm (described below) and then ran a systematic search of the best combination of model hyperparameters (i.e., scikit-learn GridSearchCV [29]) to get the tuning parameters on the training model. We used five-fold cross-validation on training data with the identified hyperparameters to evaluate which model was best to train the model [30]. We used a split of 80/20 in each fold and saved the evaluation results on the validation sample. We analyzed the mean validation score from each fold. Based on the cross-validation evaluation result, eXtreme Gradient Boosting (XGBoost) achieved the best results across all cognitive domains. We then trained the XGBoost model on all training samples and predicted the results on test/hold-out samples.

Feature Extraction. We used a combination of the following algorithms for feature selection: random forest feature ranking, linear model feature ranking using linear regression, recursive feature elimination using linear regression, and stability selection via randomized LASSO. The final importance score was based on the mean of the results of all four methods.

Model Development. For the classification task, we defined cognitive variables capturing eight cognitive domains as outcomes to predict (motor coordination normalization, emotion bias normalization, processing speed normalization, sustained attention normalization, controlled attention normalization, flexibility normalization, inhibition normalization, and executive function normalization). We fitted data into various traditional ML algorithms like support vector machines, random forest, XGBoost, Gradient Boosting Machine, AdaBoost, linear, ridge, elastic net, and LASSO regression using a grid search for hyperparameter tuning of the hyperparameter learning rate, maximum depth, maximum child weight, number of estimators, alpha, lambda, and subsample and compared results between them [31].

Model Evaluation. We used well-established metrics to measure the accuracy of the model’s prediction of our continuous variables $R^2$, mean absolute error (MAE), root-mean-squared error (RMSE), and mean absolute percentage error (MAPE) [31]. The MAE captures the average of the absolute differences between prediction and actual observation on the test fold across all cross-validation iterations where all individual differences have equal weight. The RMSE is similar, but since it uses the square root, it penalizes larger errors stronger than smaller errors. If all errors in the test set fold have the same magnitude, MAE and RMSE are the same. The closer MAE and RMSE are to zero the better. The error is given in units of the variable of interest for both RMSE and MAE so that it has the same interpretation as the outcome measures. The RMSE can be interpreted as the average distance between the observed values and the predictions. The MAE is scale-independent and has the advantage of being easily interpreted as the percentage of the prediction error and allows to compare the error across outcome scales. $R^2$ is useful to better understand how much variance in the outcome measure can be explained by the model [32]. The values of $R^2$ range from 0 to 1 and we multiplied them by 100 to facilitate the interpretation as percentages. The “optimal” model was selected by choosing the model with the lowest MAPE score (if the MAPE score was the same, we used the $R^2$ score). Balancing the cross-validated $R^2$ against the lowest cross-validated loss score was utilized to guard against overfitting due to the model complexity of the applied algorithms and the large and high-dimensional feature sets [31].
Results

We extracted 247 voice and audio features in 81 trauma survivors (34 [42.5%] women; age 37.86 ± 13.99 years; 20 [25%] Hispanic). Descriptive statistics for each domain of the cognitive performance are presented in online supplementary Table 1.

Regarding predictive model performance, we analyzed the results for various ML models, but XGBoost achieved the highest predictive score for the majority of outcomes, balancing the lowest MAPE score against the highest $R^2$ score against the lowest selected loss score. Online supplementary Tables 2–9 present the predictive matrix for various ML models. These conditions were met for model selection for all cognitive variables (Fig. 1).

The XGBoost algorithm achieved high discriminatory accuracy for predicting the different cognitive domains using cross-validation ($R^2$ ranged from 42 to 69%, MAPE ranged from 0.024 to 0.076; Fig. 1). These results were stable on the hold-out test set (Table 2).

All four domains (face, voice, speech content, and head movement) contributed significantly to the prediction (Fig. 2).

Ranking of the Features for Predictive Values

The feature importance graphs show variable feature importance using random forest, linear model, recursive feature, and stability selection. All four domains (face, voice, speech content, and movement) were ranked highly under the most important predictors (Fig. 2).

Based on the feature engineering step, we selected the top 15 features for each neurocognitive symptom prediction. The top features from each category are ranked in Figure 2.

Discussion

Accessible and minimally burdensome methods for frequent measurement of cognitive functions are needed to support research and clinical decision-making in different clinical contexts. Our results indicate that facial, movement, and voice features that are measurable in an automated manner using free speech interviews can be utilized to assess multiple domains of cognitive functioning simultaneously. Cross-validated results demonstrate accuracy in the measurement of all cognitive domains consistent with high test-retest reliability on traditional cognitive assessments [33]. Our results show the technical feasibility and that our digital biomarkers were able to predict cognitive performance with high accuracy. All interviews were successfully recorded. Data processing and prediction of cognitive performance was possible for all subjects. This innovative approach to digital phenotyping provides an opportunity to greatly scale assessments that are relevant for risk assessment and treatment across multiple psychiatric and neurological patient populations [34, 35].

Our ML algorithm considers possible nonlinear relationships of the included features in a data-driven way and incorporates regularization to reduce the risk of overfitting [36]. The feature importance represents the ranking of the most important features that are predicting each cognitive domain. The predictive accuracy is achieved by the combination of all features together. The results show that the combination of all the different feature groups together (i.e., facial features, voice prosody, movement, and speech content) is relevant for predicting each cognitive domain. This is consistent with previous results demonstrating that the combination of digital measurements of human behavior can serve as proxy digital biomarkers of multiple domains of cognitive functioning [11]. One limitation is the difficulty to fully interpret ML results. This limitation arises because the predictive performance is achieved by mathematically complex models, making it challenging to determine how the selected features relate to each other in the high-dimensional feature space of the constructed model. The interpretation of the variable importance ranking as univariate associations is limited and should be interpreted by acknowledging that the prediction is always an achievement of all variables together. However, we observe that the features selected as important indicators largely match with previous reports from the literature on the relevance of motor functioning and emotions on cognitive performance [37–39]. For example, features that were highly

### Table 2. Results on the hold-out test set (XGBoost algorithm)

|                           | $R^2$ | MAE   | RMSE  | MAPE, % |
|---------------------------|-------|-------|-------|---------|
| Motor coordination        | 0.520 | 0.078 | 0.109 | 0.056   |
| Emotion bias              | 0.524 | 0.097 | 0.163 | 0.080   |
| Processing speed          | 0.420 | 0.069 | 0.100 | 0.044   |
| Sustained attention       | 0.513 | 0.067 | 0.084 | 0.045   |
| Controlled attention      | 0.440 | 0.123 | 0.149 | 0.075   |
| Flexibility               | 0.426 | 0.087 | 0.134 | 0.057   |
| Inhibition                | 0.639 | 0.103 | 0.121 | 0.064   |
| Executive function        | 0.634 | 0.034 | 0.044 | 0.025   |

MAE, mean absolute error; MAPE, mean absolute percentage error; RMSE, root-mean-squared error; XGBoost, eXtreme Gradient Boosting.
Fig. 2. Variable importance.
ranked as relevant to motor coordination represented a large set of facial expressivity, head movement, and vocal features, all indicative of overall movement rates [37, 38]. Emotional bias was highly predicted by the presence of negative emotion and its intensity [37, 38].

The current work presents limitations, including the lack of replication in an independent sample and the relatively small sample. Further, the study drew from a community sample that had experienced a traumatic event rather than capturing the full range of variability in the cognitive performance of the general population. These results should be replicated and extended to other populations to allow for the development of clinically meaningful scores that can be derived from this modeling approach.

Importantly, these results indicate that the development of both models and technology to assess cognitive functioning based on a semi-structured snapshot of an individual’s conversation opens new avenues for the assessment of cognitive functioning at scale. Indeed, such methodologies will improve cognitive measurement in multiple ways and allow them to assess different cognitive domains simultaneously. If replicated in a larger and more heterogenous sample, the approach is promising as a validated alternative to clinician-administered neurocognitive test batteries, with high potential to develop an objective, flexible, economical, and ecologically valid digital biomarker. This is due to several advantages of our approach over clinician-administered testing. First, our ML-based digital phenotyping approach does not require data to be obtained using a rigidly structured clinical interview, which simplifies the assessment under diverse clinical context and practical constraints. Second, the approach eliminates the need for a test for each domain of cognitive functioning and provides predictions of all domains simultaneously. Third, digital phenotyping will allow a lightweight cognitive assessment to be deployed remotely, and this can lead to reducing cost but also to broader applications, including in situations of emergencies or natural disasters such as the COVID-19 pandemic where personal contact is considered unsafe. Fourth, because of the lack of a formal “test,” the evaluations can be performed theoretically at any frequency without practice effects. Future studies should examine this directly and include a study design using repeated measurements. Finally, such methods greatly increase the scalability and reduce the cost of cognitive assessment. These factors are highly relevant as individuals at risk for cognitive impairment are challenged both in their functioning and their resources [1–5].

Statement of Ethics

The authors assert that all procedures contributing to this work complied with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All participants signed informed consent and all procedures were reviewed, approved, and monitored by the New York University Institutional Review Board.

Conflict of Interest Statement

I.R. Galatzer-Levy and V. Yadav receive salary from AiCure but have no stocks. All other authors declare no potential conflict of interest that is relevant to this article.

Funding Sources

K. Schultebraucks was supported by the German Research Foundation (SCHU 3259/1-1). The study was also supported by K01MH102415 (I.R. Galatzer-Levy).

Author Contributions

K. Schultebraucks and I.R. Galatzer-Levy substantially contributed to the design of the study and the development of the study concept and developed the data analytical plan. V. Yadav performed data analysis. K. Schultebraucks and I.R. Galatzer-Levy provided supervision. K. Schultebraucks wrote the first draft of the manuscript and all co-authors reviewed and revised the manuscript. All authors approved the version of the manuscript to be published.

References

1. Brandes D, Ben-Schachar G, Gilboa A, Bonne O, Freedman S, Shalev AY. PTSD symptoms and cognitive performance in recent trauma survivors. Psychiatry Res. 2002 Jul; 110(3): 231–8.
2. Mueller-Pfeiffer C, Martin-Sobelch C, Blair JR, Carnier A, Kaiser N, Rufer M, et al. Impact of emotion on cognition in trauma survivors: what is the role of posttraumatic stress disorder? J Affect Disord. 2010 Oct; 126(1–2): 287–92.
3. Samuelson K, Newman J, Abu AD, Qian M, Li M, Schultebraucks K, et al. Predeployment neurocognitive functioning predicts postdeployment posttraumatic stress in Army personnel. Neuropsychology. 2020 Mar;34(3): 276–87.
4. Hammar A, Årdal G. Cognitive functioning in major depression – a summary. Front Hum Neurosci. 2009 Sep;3:26.
5 Kemp AH, Gordon E, Rush AJ, Williams LM. Improving the prediction of treatment response in depression: integration of clinical, cognitive, psychophysiological, neuroimaging, and genetic measures. CNS Spectr. 2008 Dec;13(12):1066–86.
6 Aupperle RL, Melrose AJ, Stein MB, Paulus MP. Executive function and PTSD: disengaging from trauma. *Neuropsycharmacology*. 2012 Feb;62(2):686–94.
7 Ben-Zion Z, Fine NB, Keynan NJ, Admon R, Green N, Halevi M, et al. Cognitive flexibility predicts PTSD symptoms: observational and interventional studies. *Front Psychiatry*. 2018 Oct;9:477.
8 Southwick SM, Charney DS. The science of resilience: implications for the prevention and treatment of depression. *Science*. 2012 Oct;338(6103):79–82.
9 Carmi L, Schultebrack K, Galatzer-Levy I. Identification, prediction, and intervention via remote digital technology: digital phenotyping & deployment of clinical interventions following terror and mass casualty events. In: Vermueten E, Frankova I, Carmi L, Chaban O, Zohar J, editors. Management of terrorism induced stress – guideline for the golden hours. Vol 148. Amsterdam: IOS Press BV; 2020. p. 175–81.
10 Insel TR. Digital phenotyping: technology for a new science of behavior. *JAMA*. 2017 Oct; 318(13):1215–6.
11 Dagum P. Digital biomarkers of cognitive function. *NPJ Digit Med*. 2018 Mar;1(1):10.
12 Silverstein SM, Berten S, Olson P, Paul R, Williams LM, Cooper N, et al. Development and validation of a World-Wide-Web-based neurocognitive assessment battery: WebNeuro. *Behav Res Methods*. 2007 Nov;39(4):940–9.
13 Martinez-Martín N, Insel TR, Dagum P, Greely HT, Cho MK. Data mining for health: staking out the ethical territory of digital phenotyping. *NPJ Digit Med*. 2018;1(1):68.
14 Poria S, Cambria E, Bajpai R, Hussain A. A validated predictive algorithm of post-traumatic stress course following emergency department admission after a traumatic stressor. *Nat Med*. 2020 Jul;26(7):1084–8.
15 Yang Y, Fairbairn C, Chaban O, Carmi L, Shalev AY, Michopoulos V, Grudzen CR, Shin SM, Stevens JS, et al. A validated predictive algorithm of post-traumatic stress course following emergency department admission after a traumatic stressor. *Nat Med*. 2020 Jul;26(7):1084–8.
16 He Q, Veldkamp BP, de Vries T. Screening for posttraumatic stress disorder using verbal features in self narratives: a text mining approach. *Psychiatry Res*. 2012 Aug;198(3):441–7.