Real-Time Assessment of Stress and Stress Response Using Digital Phenotyping: A Study Protocol

Stephan T. Egger [1,2]*, Marius Knorr [1], Julio Bobes [2], Abraham Bernstein [3], Erich Seifritz [1] and Stefan Vetter [1]

[1] Department of Psychiatry, Psychotherapy and Psychosomatics, Faculty of Medicine, Psychiatric University Hospital of Zurich, University of Zurich, Zurich, Switzerland, [2] Department of Psychiatry, Faculty of Medicine, University of Oviedo, CIBERSAM, Oviedo, Spain, [3] Department of Informatics, University of Zurich, Zurich, Switzerland

Background: Stress is a complex phenomenon that may have a negative influence on health and well-being; consequently, it plays a pivotal role in mental health. Although the incidence of mental disorders has been continuously rising, development of prevention and treatment methods has been rather slow. Through the ubiquitous presence of smartphones and wearable devices, people can monitor stress parameters in everyday life. However, the reliability and validity of such monitoring are still unsatisfactory.

Methods: The aim of this trial is to find a relationship between psychological stress and saliva cortisol levels on the one hand and physiological parameters measured by smartphones in combination with a commercially available wearable device on the other. Participants include cohorts of individuals with and without a psychiatric disorder. The study is conducted in two settings: one naturalistic and one a controlled laboratory environment, combining ecological momentary assessment (EMA) and digital phenotyping (DP). EMA is used for the assessment of challenging and stressful situations coincidentally happening during a whole observation week. DP is used during a controlled stress situation with the Trier Social Stress Test (TSST) as a standardized psychobiological paradigm. Initially, participants undergo a complete psychological screening and profiling using a standardized psychometric test battery. EMA uses a smartphone application, and the participants keep a diary about their daily routine, activities, well-being, sleep, and difficult and stressful situations they may encounter. DP is conducted through wearable devices able to continuously monitor physiological parameters (i.e., heart rate, heart rate variability, skin conductivity, temperature, movement and acceleration). Additionally, saliva cortisol samples are repeatedly taken. The TSST is conducted with continuous measurement of the same parameters measured during the EMA.

Discussion: We aim to identify valid and reliable digital biomarkers for stress and stress reactions. Furthermore, we expect to find a way of early detection of psychological stress in order to evolve new opportunities for interventions reducing stress. That may allow us to find new ways of treating and preventing mental disorders.

Trial Registration: The competing ethics committee of the Canton of Zurich, Switzerland, approved the study protocol V05.1 May 28, 2019 [BASEC: 2019-00814]; the trial was registered at ClinicalTrials.gov [NCT04100213] on September 19, 2019.

Keywords: stress, cortisol, trier social stress test, digital phenotyping, ecological moment assessment

BACKGROUND

Stress is a complex natural phenomenon, broadly defined as "the non-specific response of the body to any demand" (1). Oversimplified, this response can be divided into two components: the physiological reaction on the one hand and the subjective experience on the other (2, 3). Physiological stress causes the liberation of hormones (mainly adrenalin and cortisol) and the activation of the autonomic nervous system (1–6), resulting in changes in several physiological variables, including heart rate, heart rate variability, respiratory rate, skin conductance, and temperature (3–5, 7). The response on the behavioral level varies greatly; broadly, it may be conceived as a freeze, flight, fight, fright, or faint response (8).

So far, many studies demonstrate the negative influence of psychological stress on health and well-being (7) with several somatic and even some psychiatric disorders etiologically linked to stress (6, 9). Furthermore, mental disorders are generally conceived as harmful dysfunctions of psychological coping and adaptation mechanisms (10). For nearly three decades, the incidence of mental disorders has been continuously rising worldwide (11, 12), and this consistently accounts for a substantial proportion of social costs and the burden of disease (11, 12). The increment of psychiatric disorders has been attributed in Western societies to the rise in stress levels. The development of methods to either prevent psychiatric disorders or significantly improve their outcome has, by contrast, been slow (12).

Digital technology and information sciences are expected to profoundly change the way we understand and approach mental health (13), for example, the ubiquitous presence of smartphones (13) and the increasing availability and affordability of wearable devices capable of measuring bodily functions (14). Digital phenotyping (DP) seeks to find digital biomarkers, particularly for cognition, stress, and behavior (13, 15–19), by assessing smartphone interaction and voice and speech features, together with monitoring movement and physiological parameters (20, 21). However, from current studies (15, 17), together with earlier psychological studies (22, 23), it becomes clear that a proper validation of the users' individual emotional experience is essential (13, 16, 24–26).

Through the DP of physiological and psychological stress reactions, in real-life situations and a controlled laboratory setting, in a population of healthy participants and patients with a psychiatric disorder, we expect to find reliable and valid digital

biomarkers. Therefore, we plan to conduct a psychological and physiological study, combing ecological momentary assessment (EMA) and a laboratory psychological paradigm to induce stress, namely the Trier Social Stress Test (TSST) (27).

METHODS/DESIGN

The aim of the present trial is to establish a relationship between the physiological parameters measured by commercially available wearable devices and changes in cortisol levels obtained during everyday difficult and stressful situations and a controlled stress situation. We expect to establish a valid and reliable DP for stress and stress reactions as well as for patients with a psychiatric disorder and otherwise healthy subjects.

Participants

Participants include cohorts of participants with and without a psychiatric disorder; those with a psychiatric disorder are further categorized according to diagnosis into internalizing, externalizing, or psychotic (thought) disorders. To ensure generalizability of the findings and to minimize confounders, an overall physically healthy sample is crucial. Another critical factor is hand preference because it can influence the measurement and, therefore, reduce generalizability (28); for convenience, we include only right-handed persons. The inclusion and exclusion criteria are summarized in **Table 1**, and they are determined through the collection of a complete medical (and psychiatric) history and a medical exam (**Figures 1**, **2**). All participants undergo the same procedures, regarding

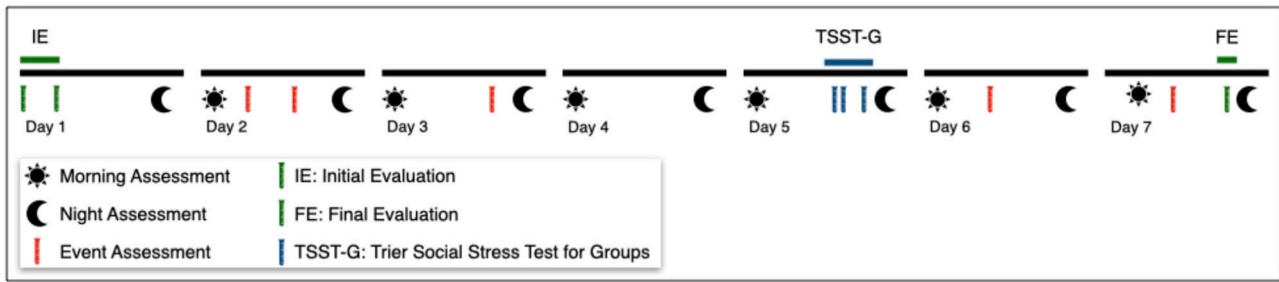**TABLE 1** | Inclusion and exclusion criteria for trial participation.

Inclusion criteria

Participants are between 18 and 65 years of age

Participants are competent to give informed consent

Right-handedness as determined by the Hand Preference Questionnaire (HPQ) (29)

German language proficiency as a native speaker or level B1 according to the Common European Framework of Reference for Languages (CEFRL) (30)

Diagnosis of a cluster C personality disorder according to ICD-10 (31); or

Diagnosis of a depressive disorder according to ICD-10 (31); or

Diagnosis of schizophrenia or schizoaffective disorder according to ICD-10 (31); or

Without a current psychiatric disorder

Exclusion criteria

Low Intelligence as confirmed by failure to complete regular compulsory education

Pregnancy or over 2 weeks delay in the menstrual cycle

Previous participation in a psychological trial involving psychosocial stress assessment

Current neurological disorder

Current cardiovascular disorder

Current respiratory disorder

Current substance use or withdrawal

Any change in medication in the previous week

FIGURE 1 | Study Outline. The initial and final evaluations (green) include a psychological test battery and the collection of cortisol samples. Well-being and basal cortisol levels are assessed daily at fixed time frames (only morning and night assessment shown). During the TSST-G (blue) DP, stress and cortisol levels are conducted. Coincidentally experienced challenging of stressful situations/events (red) are assessed shortly after they occur (shown for illustrative purposes only). Physiological parameters are continuously assessed through wearable devices.

| | | Study Procedures | | | | | | | |
| | | EMA | Digital Phenotyping | | | | | EMA | |
	Enrolment	Day 1-5	$t_{i-2/-1}$	t_{i0}	t_{i1}	t_{i2}	$t_{i3/9}$	Day 6-7	Close-Out
Eligibility Screen	X								
Informed Consent	X								
Diagnostic Group	X								
Intervention									
TSST-G				●———————●					
Tension Release			●———●			●———●			
Assessments									
Descriptive Data	X								
Medical History	X								
Medical Exam	X								
Psychometric Tests	X								X
Wake/Sleep		X						X	X
Daily Activities		X	X					X	
Stress Perception		X	X	X	X	X	X	X	
Relax Perception		X					X		
Physiological Data		X	X	X	X	X	X	X	
Salivary Cortisol		X	X	X	X	X	X	X	

FIGURE 2 | SPIRIT Study Schedule (EMA, Ecological Momentary Assessment; TSST-G, Trier Social Stress Test for Groups). The bar denotes the different parts of the intervention (compare **Figure 3**). $t_{i-2/-1}$ briefing and tension release exercise previous to the TSST-G; t_{i1} TSST-G Speech; t_{i2} TSST-G Math; $t_{i3/9}$ TSST-G debriefing and tension release.

psychometric screening and profiling, EMA, DP, and the TSST for groups.

Psychometric Measurements

All participants (regarding their psychiatric condition) undergo a full psychological screening and profiling with a standardized psychometric test battery, including self-administered and observational instruments. Raters of the following are psychiatry residents or clinical psychologists. They are trained in specific workshops on the use and objectives of the measures used in the study. The workshops follow a standardized schedule, using case vignettes and video examples. Refresher training sessions are provided regularly with trainers available for consultation at any time. The psychometric instruments included in the test battery are summarized in **Table 2**.

EMA and DP

The phenomenological assessment usually relies on a first-person narrative account collected at research or clinical visits. Self-reports, however, are known to sometimes be inaccurate for several reasons, for example, that events fade from memory over time. In contrast, EMA allows the timely record of a

TABLE 2 | Psychometric measurement instruments included in the Test Battery used in the study.

Instrument	Description
Brief Neurocognitive Assessment (*BNA*)	The Brief Neurocognitive Assessment (*BNA*) was developed as a brief, easily applicable and reliable tool to evaluate global neurocognition and impairment, primarily in patients with a diagnosis of schizophrenia (32)
Brief Symptom Inventory (*BSI*)	The Brief Symptom Inventory (*BSI*) (33) is a self-administered questionnaire assessing psychological distress; it can be used either for screening or outcome evaluation
Clinical Global Impression (*CGI*)	The Clinical Global Impression (*CGI*) scale was initially introduced in psychopharmacological trials (34). It is a brief, easy-to-use, and pragmatic tool for the assessment of psychiatric illness severity and changes over time (35)
Global Assessment of Functioning (*GAF*)	The Global Assessment of Functioning (*GAF*) is widely used in psychiatric research. It is a single-item observer-rated scale of overall functioning on a continuum from mental health to mental illness (36)
Hamilton Anxiety Rating (*HAM-A*)	The Hamilton Anxiety Rating Scale (*HAM-A*) comprises 14 items and provides an overall measure of anxiety, including psychological, cognitive, and somatic symptoms (35, 37). The scale can be used to measure anxiety in various psychiatric conditions (35)
Hamilton Depression Scale (*HAM-D*)	The Hamilton Depression Rating Scale (*HAM-D*) is a checklist of 21 items designed to measure the severity of depression (35, 38). Besides depression, it has also been used to measure depressive symptoms in other disorders (35)
Health of the Nation Outcome Scales (*HoNOS*)	The Health of the Nation Outcome Scales (*HoNOS*) is an observer-rated scale to assess the severity of a psychiatric disorder in four dimensions: behavior, impairment, symptoms, and social problems (39, 40)
Insecurity Questionnaire (*IQ-24*)	The Insecurity Questionnaire [German: Unsicherheitsfragebogen] (*IQ-24*) is a self-administered questionnaire with 24 items developed to assess insecurity (41)
Mini ICF- APP (*mICF*)	The Mini ICF-APP (*mICF*) is a short observer-rated scale to assess the level of functioning and capacity. It is easy to use and possesses good psychometric properties (42, 43)
Mini-International Neuropsychiatric Interview (*MINI*)	The Mini-International Neuropsychiatric Interview (MINI) (44) is a structured diagnostic interview. It was designed as a quick but accurate structured psychiatric diagnostic interview (according to DSM-IV and ICD-10) for clinical trials and epidemiology studies
Montreal Cognitive Assessment (*MoCA*)	The Montreal Cognitive Assessment (MoCA) (45, 46) is a brief cognitive screening tool to detect mild cognitive impairment in patients performing in the normal range on the Mini-Mental Status Exam
Positive and Negative Syndrome Scale (*PANSS*)	The Positive and Negative Syndrome Scale (*PANSS*) is a semistructured interview designed to measure the severity of psychopathology in patients with a psychotic disorder: mainly schizophrenia and schizoaffective disorder (35, 47). The PANSS measures symptoms in three domains: positive, negative, and general symptoms
Protocol for Sleep Examination (*PSE*)	The Protocol for Sleep Examination [German: Abend/Morgenprotokolle für die Schlafuntersuchung] (*PSE*) was developed to assess the subjective dimension of sleep, daily activities, and distress. It is divided into morning and night (bedtime) sections (48)

(Continued)

TABLE 2 | Continued

Instrument	Description
Short Stress Questionnaire (*SSQ*)	The Short Stress Questionnaire [German: Kurzfragebogen zur aktuellen Beanspruchung] (*SSQ*) was developed to assess subjective levels of tension or stress associated with a current task, situation, or experience (49)
Toronto Alexithymia Questionnaire (*TAQ*)	The Toronto Alexithymia Questionnaire (*TAQ*) (50, 51) was developed to assess difficulties identifying subjective emotional feelings, distinguishing between feelings and the bodily sensations of emotional arousal and difficulty describing feelings to other people (51, 52)
Yale-Brown Obsessive-Compulsive Scale (*Y-BOCS*)	The Yale-Brown Obsessive Compulsive Scale (*Y-BOCS*) was developed to measure the severity of obsessive-compulsive symptoms; these are rated in terms of time spent on such activities, interference with functioning, distress, resistance, and control (35, 53)
Young Mania Rating Scale (*YMRS*)	The Young Mania Rating Scale (*YMRS*) is an observer-rated checklist, measuring manic symptoms to quantify the severity and the effect of treatment (35, 54). It includes the core symptoms of mania occurring in both mild and severe illness.

person's experience and behavior in the natural environment, thus, increasing the validity and allowing the inference of factors influencing behavior and experience. EMA is a long-known methodology in psychological and anthropological research, usually with the use of dairies or logbooks. The appearance of smartphones and wearable devices facilitates the implementation of EMA studies (55, 56).

EMA is conducted over a whole week using a custom smartphone application and two wearable devices. Through the smartphone application, participants are able to evaluate their daily activities and sleep. In addition, the application prompts the participants once or twice a day about their current activity. Participants are able to log any stressful and challenging situation. Participants have to answer a short questionnaire regarding their current activity, well-being, and stress level (see **Figures 1**, **2**). Through two commercially wearable devices (Vívosmart® wristband and Everion® armband), several physiological parameters are continuously monitored and recorded, including heart rate, skin conductance, temperature, movement, and acceleration (see **Table 3**). We included two devices in order to allow for comparison and generalizability of the results, especially taking into account possible flaws in the use and the measurement quality of the devices (57).

Cortisol secretion follows a circadian rhythm, usually with a peak in the morning and slowly declining throughout the day with variations from day to day and individual to individual (58). Therefore, for proper validation and interpretation, regular measurements of cortisol levels are necessary (59, 60). Participants collect a saliva sample four times a day (morning, midday, afternoon, and night); after experiencing a difficult or stressful situation and at random once or twice a day. Saliva samples are picked up and sent once a day (at night) to the laboratory for the quantification of cortisol levels, and after analysis, samples are destroyed.

TABLE 3 | Digital parameters for Ecological Momentary Assessment (EMA) and Digital Phenotyping (DP).

Parameter	Smartphone	Everion®	Vívosmart®
Heart rate		X	X
Heart rate variability		X	X
Blood pulse wave		X	
Respiratory rate		X	
Oxygenation		X	
Skin blood perfusion		X	
Skin temperature		X	
Electrodermal activity		X	
Activity/movement	X	X	X
Distance	X		
Altitude	X		X
Location	X		

Trier Social Stress Test for Groups (TSST-G)

The TSST is an extensively used and well-validated psychological paradigm to induce psychobiological stress in laboratory settings (61–64) with a significant association with an acute stress response in real life (62, 65, 66). The TSST has been modified in order to be conducted in groups; in our current study, we include three to five participants (from the same diagnostic group), using TSTT-G procedures analogous to previous studies (67–69). The TSST-G consists of three phases: a briefing, the psychological test itself, and a debriefing. The phases last 40, 20, and 60 min, respectively (**Figures 2**, **3**). An experienced psychotherapist conducts the briefing and debriefing of the TSST-G. The TSST-G itself is conducted by personnel unknown to the participants. During the TSST-G, saliva samples are obtained at regular intervals and cortisol levels measured.

Each participant undergoes an individual briefing phase. Participants are required to prepare a speech for a job application. After a few minutes, participants are accompanied into the test room and are seated next to each other, separated by partitions in order to avoid eye contact. They are told that an expert committee will conduct an analysis of their performance and that they will be videorecorded (no actual recording is performed) for further analysis. The participants present their speech (2–3 min each) in a previously set random order. Next, the participants conduct a subtraction task (for 2 min) as quickly and as accurately as possible. If participants make a mistake, they are asked to start over again. The order of participation once again is random. Once the last participant has completed the task, the committee leaves the room. Participants are accompanied back to the preparation room, where they are debriefed and may engage in any relaxing activity for 60 min.
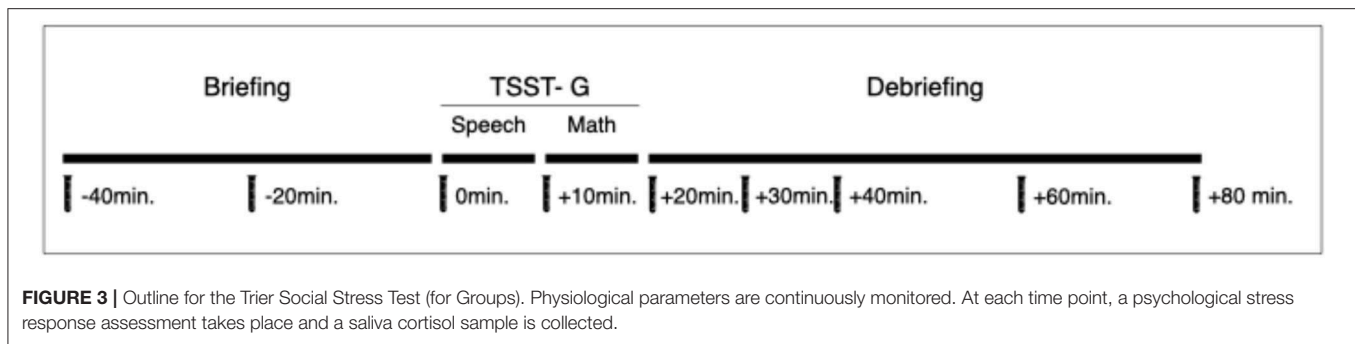
Sample Size Calculation and Statistical Analysis

Previous research has consistently shown that the TSST significantly increases the cortisol levels with moderate effect sizes regarding baseline (63). Therefore, we expect a low to moderate effect size in cortisol through the TSST-G in our study. We calculate our required sample size using G*Power 3.1 (70) (ANOVA: repeated measures, within and between factors; effect size $f = 0.4$; $\alpha = 0.05$; power $= 0.8$; number of groups $= 4$, number of measures $= 9$, nonsphericity correction $= 0.125$). Based on that calculation, at least 24 participants per group are required to detect moderate-sized differences: to improve capacity, we include at least 30 participants in each group. Only data sets of participants who complete the intervention are considered (completed TSST-G and at least 70% completion of the EMA); therefore, recruitment continues until the number of participants for each group is reached. Already enrolled participants are able to complete the study.

The primary analysis is conducted with complete cases only; dropouts are replaced by recruiting new subjects. Secondary analysis includes incomplete cases and dropouts. If a participant withdraws from the study, his or her data is anonymized and his or her name is deleted permanently from all study records. Unless otherwise stated, his or her remaining data is used in the secondary analysis. Data analysis does not pursue hypothesis testing; through the statistic scrutiny of the data, we aspire to gain a better understanding of the possibilities offered by wearable devices for the assessment of stress and stress reactions and finding digital biomarkers. Accordingly, the findings of the study serve for the formulation of hypothesis and hypothesis testing in future studies.

The demographic and clinical characteristics of the sample are compared at baseline using an ANOVA, excluding gender, which is analyzed using the chi-square test. Repeated-measures multivariate analyses of variance (MANOVAs) are used to assess changes in symptomatology, functionality, cognition, and physiological parameters. To infer differences in stress reactions according to the subjective experience and clinical characteristics, we use a multivariate regression analysis as well as time series analysis. To avoid inflation of type II errors, we apply a Bonferroni correction for multiple comparisons. The significance threshold is set at 0.05. Cohen's d is calculated to determine the effect size (71). Multiple and logistic analyses as well as time series analysis is performed. Due to the complexity of the data, with a large number of variables and potential confounders, a machine learning algorithm is used to detect complex relationships between the stress, psychopathology, and physiological measures (72).

For each wearable device, machine learning is conducted stepwise, using a supervised learning approach at first and a deep learning approach at last. For analysis, three separate data sets are created. The first data set comprises the measures collected during the TSST-G with the speech and math as stress events and the briefing and debriefing as relaxing events. This data set is subdivided into two sets: one for training the model and one for testing. One stress and one relaxing event are randomly assigned to either one of the data sets. The second data set consists of the three full-day measurements selected at random: two from the days previous and one from the days after the TSST-G. The second data set is used for the deep learning algorithm for the detection of stress and relaxation. The third and final data set comprises the remaining days: two previous and one

FIGURE 3 | Outline for the Trier Social Stress Test (for Groups). Physiological parameters are continuously monitored. At each time point, a psychological stress response assessment takes place and a saliva cortisol sample is collected.

after the TSST-G. This data set will be used for testing the obtained models.

Quality of Data, Missing Data

The design of our trial, with the preparation of the probands and instruments, allows us to ensure that measurements obtained during the TSST-G have high quality with a low artifact rate. Due to the complexity and duration of the remaining intervention, we cannot rule out that all the measurements obtained will reach a high-quality threshold. The use of a device (Everion®) with a high measurement quality as well as its placement (57) should increase the quality of the measurements. Missing measurements and artifacts from the digital devices are not replaced. Missing items in the different psychometric instruments are replaced according to the rules and conventions for each instrument.

TRIAL STATUS

The competing ethics committee of the Canton of Zurich, Switzerland, approved the study protocol V05.1- May 28, 2019 [BASEC: 2019-00814]; the trial was registered at ClinicalTrials.gov [NCT04100213] on September 19, 2019. Recruitment starts in Fall/Winter 2020. We expect to recruit the whole sample in 9 to 12 months from the first enrollment.

DISCUSSION

Stress is a known risk factor for several, if not all, psychiatric disorders. However, the perception and reaction to stress show considerable variability among the general population and even more among those suffering from a psychiatric disorder. Healthy subjects are more or less consciously aware of stress and potentially stressful situations and, therefore, able to adjust and modify their behaviors in order to master life's challenges. Patients with a psychiatric disorder, conversely, have a disrupted perception, awareness, and reaction to stress (2, 73, 74), hampering them in adapting and coping with everyday demands. Stress has, therefore, become a major target of lifestyle and well-being and psychiatric prevention and treatment research with several interventions focusing on stress awareness and management.

The use of smartphones and wearable devices nowadays is ubiquitous with a significant increase in their application

to monitor psychological well-being and stress. Uncountable digital services are claiming to appraise and improve physical and psychological well-being (26, 75). However, despite gaining popularity, their use remains controversial. Users frequently experience deception (76, 77), generally due to privacy and confidentiality issues (25, 76, 78, 79) but also inaccurate feedback or even dangerous advice (25, 26, 79).

There is still a lack of guidance in the use of such devices in general and in psychiatry in particular with guidelines and legal regulations that are still emerging (61, 80). From the available services, only a tiny fraction has been validated adequately in controlled studies (24, 26, 81). Persons with a psychiatric disorder are under-represented in current studies, reducing the use and applicability of such devices in psychiatric settings. Their reckless use may be detrimental, dangerous, or even harmful (81, 82).

From current digital trials (15, 17) and earlier psychological studies (22, 23), it is clear that proper validation and fitting to the users' individual emotional experience is required (13, 16, 24–26). We consider it essential to assess the individual stress and stress reactions in everyday situations and a controlled laboratory setting. The TSST (27) is an extensively used and well-validated psychological paradigm to induce psychobiological stress (61–64) with a significant association with acute stress response in real life (62, 65, 66).

In order to establish valid and reliable digital biomarkers, the study population is crucial (16, 20). Psychiatric diagnoses have overlapping symptoms and high psychiatric comorbidity (74, 83), making it challenging to form homogeneous groups. Therefore, in our study, we aim to establish a complete psychological profile (beyond individual diagnoses) of the participants with a transdiagnostic test battery, including the assessment of threshold and subthreshold psychiatric symptoms. Likewise, we assess their psychosocial functioning, well-being, level of stress, and coping with the challenges of daily life. Cortisol release shows variations between and within individuals (58).

The regular sampling of cortisol in saliva during a whole week allows us to establish the cortisol secretion profile for each participant; the TSST-G allows us to establish the cortisol release during a standardized, controlled, and validated psychobiological stress paradigm, therefore, giving us a "fisheye perspective" on stress and stress reaction. We expect that day-to-day situations experienced as challenging and stressful enact a similar cortisol release and physiological response as the TSST-G.

We anticipate that well-being and certain psychopathological states modify the individual's self-awareness and, therefore, the perception and reaction of challenging and stressful situations. The combination of psychopathological profiling, assessment of the subjective stress experience, physiological monitoring, and psychological observation during everyday life and under controlled and standardized laboratory conditions, however, provides a panoramic view, which, in turn, allows us to determine reliable and valid digital biomarkers.

The digital biomarkers we expect to find have the potential to facilitate self-monitoring of stress as well to serve as part of our diagnostic and therapeutic instruments. The use of both devices (high-quality and over-the-counter) allows inferring the suitability of this approach for daily use. The results obtained from this study serve for further hypothesis formulation and testing. Taking into account the complexity and dynamics in the field of digital technologies, the next step for testing and validating our results should take place in the frame of a citizen science project (84), simultaneously allowing the dissemination and improvement of the results of this study.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Canton of Zurich [BASEC: 2019-00814]. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SE: trial design, writing of the study protocol, and writing of the manuscript. MK: writing of the manuscript. JB and AB: trial design, writing of the study protocol, and correction of the manuscript. ES and SV: trial design, writing and registration of the study protocol, and correction of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Selye H. Stress and the general adaptation syndrome. *Br Med J.* (1950) 1:1383. doi: 10.1136/bmj.1.4667.1383

2. Mcewen BS. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiol Rev.* (2007) 87:873–904. doi: 10.1152/physrev.00041.2006

3. Mccarty R. Learning about stress: neural, endocrine and behavioral adaptations. *Stress.* (2016) 19:449–75. doi: 10.1080/10253890.2016.1192120

4. Draghici AE, Taylor JA. The physiological basis and measurement of heart rate variability in humans. *J Physiol Anthropol.* (2016) 35:22. doi: 10.1186/s40101-016-0113-7

5. Karemaker JM. An introduction into autonomic nervous function. *Physiol Meas.* (2017) 38:R89–118. doi: 10.1088/1361-6579/aa6782

6. Engert V, Kok BE, Puhlmann LMC, Stalder T, Kirschbaum C, Apostolakou F, et al. Exploring the multidimensional complex systems structure of the stress response and its relation to health and sleep outcomes. *Brain Behav Immun.* (2018) 73:390–402. doi: 10.1016/j.bbi.2018.05.023

7. Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. *JAMA.* (2007) 298:1685–7. doi: 10.1001/jama.298.14.1685

8. Bracha HS. Freeze, flight, fight, fright, faint: adaptationist perspectives on the acute stress response spectrum. *J CNS Spect.* (2004) 9:679–85. doi: 10.1017/S1092852900001954

9. Zannas AS, West AE. Epigenetics and the regulation of stress vulnerability and resilience. *Neuroscience.* (2014) 264:157–70. doi: 10.1016/j.neuroscience.2013.12.003

10. Wakefield JC. The concept of mental disorder. On the boundary between biological facts and social values. *Am Psychol.* (1992) 47:373–88. doi: 10.1037/0003-066X.47.3.373

11. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* (2018) 392:1789–858. doi: 10.1016/S0140-6736(18)32279-7

12. Kyu HH, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* (2018) 392:1859–922. doi: 10.1016/S0140-6736(18)32335-3

13. Insel TR. Digital phenotyping: a global tool for psychiatry. *World Psychiatry.* (2018) 17:276–7. doi: 10.1002/wps.20550

14. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. *JMIR Mhealth Uhealth.* (2016) 4:e7. doi: 10.2196/mhealth.4225

15. Dagum P. Digital biomarkers of cognitive function. *npj Digital Med.* (2018) 1:10. doi: 10.1038/s41746-018-0018-4

16. Reinertsen E, Clifford GD. A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. *Physiol Meas.* (2018) 39:05TR01. doi: 10.1088/1361-6579/aabf64

17. Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'hondt E, De Raedt W, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. *npj Digital Med.* (2018) 1:67. doi: 10.1038/s41746-018-0074-9

18. Torous J, Keshavan M. A new window into psychosis: The rise digital phenotyping, smartphone assessment, and mobile monitoring. *Schizophr Res.* (2018) 197:67–8. doi: 10.1016/j.schres.2018.01.005

19. Hirschtritt ME, Insel TR. Digital technologies in psychiatry: present and future. *Focus.* (2018) 16:251–8. doi: 10.1176/appi.focus.20180001

20. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. *Nat Biotechnol.* (2015) 33:462–3. doi: 10.1038/nbt.3223

21. Insel TR. Digital phenotyping: technology for a new science of behavior. *JAMA.* (2017) 318:1215–6. doi: 10.1001/jama.2017.11295

22. Lesser IM. A review of the alexithymia concept. *Psychosom Med.* (1981) 43:531–43. doi: 10.1097/00006842-198112000-00009

23. Larsen JK, Brand N, Bermond B, Hijman R. Cognitive and emotional characteristics of alexithymia: a review of neurobiological studies. *J Psychosom Res.* (2003) 54:533–41. doi: 10.1016/S0022-3999(02)00466-X

24. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR Ment Health.* (2016) 3:e7. doi: 10.2196/mental.4984

25. Roberts LW, Torous J. Preparing residents and fellows to address ethical issues in the use of mobile technologies in clinical psychiatry. *Acad Psychiatry.* (2017) 41:132–4. doi: 10.1007/s40596-016-0594-z

26. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. *Front Physiol.* (2018) 9:743. doi: 10.3389/fphys.2018.00743

27. Kirschbaum C, Pirke KM, Hellhammer DH. The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology.* (1993) 28:76–81. doi: 10.1159/000119004

28. Middelkoop HA, Van Dam EM, Smilde-Van Den Doel DA, Van Dijk G. 45-hour continuous quintuple-site actimetry: relations between trunk and limb movements and effects of circadian sleep-wake rhythmicity. *Psychophysiology.* (1997) 34:199–203. doi: 10.1111/j.1469-8986.1997.tb02132.x

29. Salmaso D, Longoni AM. Problems in the assessment of hand preference. *Cortex.* (1985) 21:533–49. doi: 10.1016/S0010-9452(58)80003-9

30. Verhelst N, Van Avermaet P, Takala S, Figueras N, North B. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Strasbourg: Cambridge University Press (2009).

31. World Health Organization. *ICD-10: International Statistical Classification of Diseases and Related Health Problems.* Genève: World Health Organization (2004).

32. Fervaha G, Hill C, Agid O, Takeuchi H, Foussias G, Siddiqui I, et al. Examination of the validity of the brief neurocognitive assessment (BNA) for schizophrenia. *Schizophr Res.* (2015) 166:304–9. doi: 10.1016/j.schres.2015.05.015

33. Derogatis L, Melisaratos N. The brief symptom inventory: an introductory report. *Psychol Med.* (1983) 3:595–605. doi: 10.1017/S003329170004 8017

34. Guy W. *ECDEU Assessment Manual for Psychopharmacology.* Rockville, MD: US Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs (1976).

35. Rush A Jr, First MB, Blacker DE. *Handbook of Psychiatric Measures.* Philadelphia, PA: American Psychiatric Publishing, Inc. (2008).

36. Jean Endicott P, Robert L, Spitzer M, Joseph L, Fleiss P, Jacob Cohen P. The global assessment scale. *Arch Gen Psychiatry.* (1976) 33:766–71. doi: 10.1001/archpsyc.1976.01770060036012

37. Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol.* (1959) 32:50–5. doi: 10.1111/j.2044-8341.1959.tb00467.x

38. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* (1960) 23:56–62. doi: 10.1136/jnnp.23.1.56

39. Wing J, Beevor A, Curtis R, Park S, Hadden S, Burns A. Health of the Nation Outcome Scales (HoNOS). Research and development. *Br J Psychiatry.* (1998) 172:11–8. doi: 10.1192/bjp.172.1.11

40. Wing J, Curtis R, Beevor A. Health of the Nation Outcome Scales (HoNOS). Glossary for HoNOS score sheet. *Br J Psychiatry.* (1999) 174:432–4. doi: 10.1192/bjp.174.5.432

41. Albani C, Schmutzer G, Blaser G, Körner A, Nawroth C, Geyer M, et al. Die Entwicklung einer Kurzversion (U-Bogen-24) des Unsicherheitsfragebogens von Ullrich und Ullrich de Muynck. *Psychother Psychosomatik Med Psychol.* (2005) 56:118–27. doi: 10.1055/s-2005-915332

42. Linden M, Baron S. [The "Mini-ICF-Rating for Mental Disorders (Mini-ICF-P)". A short instrument for the assessment of disabilities in mental disorders]. *Rehabilitation.* (2005) 44:144–51. doi: 10.1055/s-2004-834786

43. Baron S, Linden M. [Analyzing the effectiveness of inpatient psychosomatic rehabilitation using the mini-ICF-APP]. *Rehabilitation.* (2009) 48:145–53. doi: 10.1055/s-0029-1220740

44. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry.* (1998) 59 (Suppl. 20):22–33. doi: 10.1037/t18597-000

45. Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* (2005) 53:695–9. doi: 10.1111/j.1532-5415.2005.53221.x

46. Damian AM, Jacobson SA, Hentz JG, Belden CM, Shill HA, Sabbagh MN, et al. The Montreal Cognitive Assessment and the mini-mental state examination as screening instruments for cognitive impairment: item analyses and threshold scores. *Dement Geriatr Cogn Disord.* (2011) 31:126–31. doi: 10.1159/000323867

47. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull.* (1987) 13:261–76. doi: 10.1093/schbul/13.2.261

48. Hoffmann RM, Müller T, Hajak G, Cassel W. Abend-Morgenprotokolle in Schlafforschung und Schlafmedizin—Ein Standardinstrument für den deutschsprachigen Raum. *Somnologie.* (1997) 1:103–9. doi: 10.1007/s11818-997-0019-z

49. Müller B, Basler H-D. Kurzfragebogen zur aktuellen *Beanspruchung: KAB.* Weinheim: Beltz (1993).

50. Kupfer J, Brosig B, Brähler E. Testing and validation of the 26-Item Toronto Alexithymia Scale in a representative population sample. *Zeitschrift Psychosomatische Med Psychother.* (2000) 46:368–84. doi: 10.13109/zptm.2000.46.4.368

51. Taylor GJ, Bagby RM, Parker JDA. The 20-item toronto alexithymia scale. *J Psychosomatic Res.* (2003) 55:277–83. doi: 10.1016/S0022-3999(02)00601-3

52. Kooiman CG, Spinhoven P, Trijsburg RW. The assessment of alexithymia: a critical review of the literature and a psychometric study of the Toronto Alexithymia Scale-20. *J Psychosom Res.* (2002) 53:1083–90. doi: 10.1016/S0022-3999(02)00348-3

53. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, et al. The Yale-Brown obsessive compulsive scale: I. Development, use, and reliability. *Arch Gen Psychiatry.* (1989) 46:1006–11. doi: 10.1001/archpsyc.1989.01810110048007

54. Young R, Biggs J, Ziegler V, Meyer D. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry.* (1978) 133:429–35. doi: 10.1192/bjp.133.5.429

55. Thiele C, Laireiter A-R, Baumann U. Diaries in clinical psychology and psychotherapy: a selective review. *Clin Psychol Psychother.* (2002) 9:1–37. doi: 10.1002/cpp.302

56. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* (2008) 4:1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415

57. Barrios L, Oldrati P, Santini S, Lutterotti A. Evaluating the accuracy of heart rate sensors based on photoplethysmography for in-the-wild analysis. In: *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare - PervasiveHealth'19.* Trento (2019).

58. Dahlgren A, Kecklund G, Theorell T, Akerstedt T. Day-to-day variation in saliva cortisol–relation with sleep, stress and self-rated health. *Biol Psychol.* (2009) 82:149–55. doi: 10.1016/j.biopsycho.2009.07.001

59. Adam EK, Hawkley LC, Kudielka BM, Cacioppo JT. Day-to-day dynamics of experience–cortisol associations in a population-based sample of older adults. *Proc Natl Acad Sci USA.* (2006) 103:17058–63. doi: 10.1073/pnas.0605053103

60. Hellhammer J, Fries E, Schweisthal OW, Schlotz W, Stone AA, Hagemann D. Several daily measurements are necessary to reliably assess the cortisol rise after awakening: state- and trait components. *Psychoneuroendocrinology.* (2007) 32:80–6. doi: 10.1016/j.psyneuen.2006.10.005

61. Birkett MA. The trier social stress test protocol for inducing psychological stress. *J Vis Exp.* (2011) 56:3238. doi: 10.3791/3238

62. Allen AP, Kennedy PJ, Dockray S, Cryan JF, Dinan TG, Clarke G. The trier social stress test: principles and practice. *Neurobiol Stress.* (2017) 6:113–26. doi: 10.1016/j.ynstr.2016.11.001

63. Goodman WK, Janson J, Wolf JM. Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. *Psychoneuroendocrinology.* (2017) 80:26–35. doi: 10.1016/j.psyneuen.2017.02.030

64. Vors O, Marqueste T, Mascret N. The trier social stress test and the trier social stress test for groups: qualitative investigations. *PLoS ONE.* (2018) 13:e0195722. doi: 10.1371/journal.pone.0195722

65. Hellhammer J, Schubert M. The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. *Psychoneuroendocrinology.* (2012) 37:119–24. doi: 10.1016/j.psyneuen.2011.05.012

66. Henze GI, Zankert S, Urschler DF, Hiltl TJ, Kudielka BM, Pruessner JC, et al. Testing the ecological validity of the Trier Social Stress Test: Association with real-life exam stress. *Psychoneuroendocrinology.* (2017) 75:52–5. doi: 10.1016/j.psyneuen.2016.10.002

67. Childs E, Vicini LM, De Wit H. Responses to the Trier Social Stress Test (TSST) in single versus grouped participants. *Psychophysiology.* (2006) 43:366–71. doi: 10.1111/j.1469-8986.2006.00414.x

68. Von Dawans B, Kirschbaum C, Heinrichs M. The trier social stress test for groups (TSST-G): a new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology.* (2011) 36:514–22. doi: 10.1016/j.psyneuen.2010.08.004

69. Boesch M, Sefidan S, Ehlert U, Annen H, Wyss T, Steptoe A, et al. Mood and autonomic responses to repeated exposure to the Trier Social Stress Test for Groups (TSST-G). *Psychoneuroendocrinology.* (2014) 43:41–51. doi: 10.1016/j.psyneuen.2014.02.003

70. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* (2007) 39:175–91. doi: 10.3758/BF03193146

71. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Academic Press (1969).

72. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med.* (2019) 380:1347–58. doi: 10.1056/NEJMra1814259

73. Cheng SC, Walsh E, Schepp KG. Vulnerability, stress, and support in the disease trajectory from prodrome to diagnosed schizophrenia: diathesis-stress-support model. *Arch Psychiatr Nurs.* (2016) 30:810–7. doi: 10.1016/j.apnu.2016.07.008

74. Caspi A, Moffitt TE. All for one and one for all: mental disorders in one dimension. *Am J Psychiatry.* (2018) 175:831–44. doi: 10.1176/appi.ajp.2018.17121383

75. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. *JMIR Mhealth Uhealth.* (2016) 4:e40. doi: 10.2196/mhealth.4461

76. Tsesis A. The right to erasure: privacy, data brokers, and the indefinite retention of data. *Wake For Lake Rev.* (2014) 49:433.

77. Commission USFT. *Lumosity to Pay $2 Million to Settle FTC Deceptive Advertising Charges for Its "Brain Training" Program.* (2016). Available online at: https://www.ftc.gov/news-events/press-releases/2016/01/lumosity-pay-2-million-settle-ftc-deceptive-advertising-charges (accessed January 1, 2019).

78. Huckvale K, Prieto JT, Tilney M, Benghozi P-J, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. *BMC Med.* (2015) 13:214. doi: 10.1186/s12916-015-0444-y

79. Torous JB, Chan SR, Gipson SYT, Kim JW, Nguyen TQ, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. *Psychiatr Serv.* (2018) 69:498–500. doi: 10.1176/appi.ps.201700423

80. Veale M, Binns R, Edwards L. Algorithms that remember: model inversion attacks and data protection law. *Philos Transact Ser A Math Phys Eng Sci.* (2018) 376:20180083. doi: 10.1098/rsta.2018.0083

81. Torous J, Haim A. Dichotomies in the development and implementation of digital mental health tools. *Psychiatr Serv.* (2018) 69:1204–6. doi: 10.1176/appi.ps.201800193

82. Anthes E. Mental health: there's an app for that. *Nat News.* (2016) 532:20. doi: 10.1038/532020a

83. Plana-Ripoll O, Pedersen CB, Holtz Y, Benros ME, Dalsgaard S, De Jonge P, et al. Exploring comorbidity within mental disorders among a danish national population. *JAMA Psychiatry.* (2019) 76:259–70. doi: 10.1001/jamapsychiatry.2018.3658

84. Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, et al. Next steps for citizen science. *Science.* (2014) 343:1436–7. doi: 10.1126/science.1251554